

July 15, 2004

Version 2 (Higgs-Dilepton meeting 07/15/04)

**$h \rightarrow \gamma\gamma$ Update. (Helping Yurii Maravin with Neural
Net / Photon ID studies)**

Alex Melnitchouk
University of Mississippi

1 Final goal and initial strategy

The goal is suppress reducible photon background (multijet, γ +jet) using Neural Net relative to Moriond 2004 $h \rightarrow \gamma\gamma$ Photon ID cuts:

- CC: $\text{HMx7} < 10$, ($0.05 < dR < 0.4$) Track Isolation < 2 GeV
- EC: $\text{HMx8} < 15$, ($0.05 < dR < 0.4$) Track Isolation < 2 GeV

Initial “MC vs data and γ vs e” strategy is to start with real data and ignore for the moment the difference in efficiency between photons and electrons, i.e.

- use $Z/\gamma^* \rightarrow ee$ data for NN signal training sample and efficiency estimate
- use Single EM data for NN background training sample and fake rate estimate
- compare EMID variable distributions for signal and background (defined above) with $Z/\gamma^* \rightarrow ee$ Monte Carlo and $h \rightarrow \gamma\gamma$ Monte Carlo but not use Monte Carlo otherwise

The strategy, choice of samples, triggers, and preliminary cuts (see next section) are motivated by Moriond 2004 $h \rightarrow \gamma\gamma$ analysis ([1],[2], and [3])

2 Samples

1. diEM/ $Z/\gamma^* \rightarrow ee$ data: $\approx 270 \text{ pb}^{-1}$ of CS2EMhighpt p14 skim (d0correct/tmb_analyze of v6)
2. Single EM data: a piece of CS1EMloose p14 skim covering both pre- and post- Fall 2003 shutdown data (d0correct/tmb_analyze of v6)
3. $Z/\gamma^* \rightarrow ee$ MC: p14 (d0correct/tmb_analyze of v6a)
4. $h \rightarrow \gamma\gamma$ MC: p14 (d0correct/tmb_analyze of v6a, Higgs mass = 80, 90, 100, 110 GeV)

The discrepancy between d0correct/tmb_analyze versions of v6 for data and v6a for MC is due to the fact that tracking information is incorrect for v6 MC. Otherwise there is no significant difference between v6 and v6a to prevent valid comparison of data and MC samples.

In each of these four samples CCEC events are selected (CC: $|\eta| < 1.05$, EC: $1.5 < |\eta| < 2.4$). CC objects from CCEC events are used for CC NN training, while EC objects from CCEC events are used for EC NN training. The above clarification is irrelevant when it comes to signal efficiency, i.e. $Z/\gamma^* \rightarrow ee$ CC electron has the same efficiency whether it comes from a CCCC or CCEC event. However, as for the fake rates (defined so that fake rate denominator is loose isolated ($\text{iso} > 0.15$) em objects rather than jets), systematic dependence on the event topology is observed. For example, EC fake rate is higher in ECEC than in CCEC (see page 9 of [1] (Figure 7)).

Kinematic selections include object $p_T > 25$ GeV and $40 \text{ GeV} < \text{Mass}(\text{object1}, \text{object2}) < 160$ GeV cuts. Objects near CC calorimeter ϕ -module boundaries are accepted. Preliminary EMID cuts are $\text{emf} > 0.9$, $\text{iso} < 0.15$.

2.1 Comment on NN signal training sample

$Z/\gamma^* \rightarrow ee$ data (NN training signal) sample is derived from diEM CS2EMhighpt sample by a track match for both objects. Track matching criterion is $P(\chi^2) > 0.01$, where χ^2 describes the quality of the match between an EM object and a track based on ϕ, η , and E/p variables. The purity of such $Z/\gamma^* \rightarrow ee$ sample exceeds 99% because electron fake rate (due to QCD) for this matching criterion is $\approx 0.01\text{-}0.05$ (see page 9 of [1] (Figure 7)).

2.2 Comment on NN background training sample

Single EM data sample is a sample that consists of events with exactly one object. It is meant to contain fake em objects coming from multijet and γ +jet events as well as real photons from γ +jet events and exclude all other potential sources of EM objects. To ensure the exclusion EM sources other than objects multijet and γ +jet processes we veto the events with more than one EM object (potential $Z/\gamma^* \rightarrow ee$ and direct diphotons), as well as events with the Missing Transverse Energy above 15 GeV (potential $W \rightarrow e\nu$). Fake purity of this sample is estimated roughly to be $>86\%$ in the mass range of $M_{jj} = 100\text{-}150$ GeV. This estimate is based on [4] (page 035004-2, Figure 1), [5] (page 2246), and [6] (page 52, Figure 39).

After the selection criteria described above are applied, we end up with the following number of events:

1. $Z/\gamma^* \rightarrow ee$ data: 6k
2. Single EM data: 50k
3. $Z/\gamma^* \rightarrow ee$ MC: 9k
4. $h \rightarrow \gamma\gamma$ MC: 5k

For the purposes of NN background training (Single EM data) we use reduced sample of 10k objects.

3 Distribution of kinematic and EMID variables

Distributions of kinematic and basic EMID variables (p_T , detector η , detector ϕ , invariant mass, EM fraction, isolation, absolute isolation, and track isolation) are shown in Figures 1 and 2 for CC and EC respectively.

Figures 3 and 4 show floor energy fractions and $r\phi$ -widths for CC and EC respectively. The EM1 floor energy fraction is higher than expected given the depth of EM floors (Table 1).

This is related to the choice of the weights in the EM1. The EM1 weights were set high deliberately to improve the energy resolution. [8]

Figures 5 and 6 show z/r-widths and number of cells that contributed to a cluster in each floor for CC and EC respectively. Cells within $dR(\eta, \phi) < 0.7$ between a cell and a cluster are counted. No cell energy threshold is applied at the analysis stage. Emreco threshold is 100(?) MeV. The widths in EM1, EM2, and EM4 floors have discontinuities while the

	EM	FH	CH
CC Depth	$2, 2, 7, 10 X_0$	$1.3, 1.0, 0.9\lambda$	3.2λ
EC Depth	$0.3, 2.6, 7.9, 9.3 X_0$	$1.2, 1.2, 1.2\lambda$	3.6λ

Table 1: The depth of the calorimeter floors in the units of *radiation length* X_0 and *absorption length* λ [7].

EM3 width is continuous. We believe this is related to geometry (the granularity of EM3 is twice as fine as that of EM1, EM2, and EM3). Figure 7 shows the number of EM1 cells inside $dR < 0.7$ vs. $r\phi$ -width. As the number of cells increases the discontinuity in the $r\phi$ -width disappears.

4 Cluster profiles

We can also try make use of a cell-level cluster information and construct variables to be used in the NN. We chop up the emcluster into $4 \times 7 = 28$ regions (4 floors, 7 concentric cones/cone shells of increasing radius in steps of 0.1). In these regions we look at number of cells, absolute energy depositions, energy fractions. We repeat these studies for the cell energy threshold of 400 MeV and no threshold (100 MeV at reco ?). The plots are shown in Figures 8- 55. Table 2 is a guide that associates Figure number with cell-info variable, cell energy threshold, EM floor number, and topology(CC or EC)

	EM1: .0, .4	EM2: .0, .4	EM3: .0, .4	EM4: .0, .4
	CC	CC	CC	CC
sum of cell energies	8, 9	10, 11	12, 13	14, 15
energy fractions	16, 17	18, 19	20, 21	22, 23
number of cells	24, 25	26, 27	28, 29	30, 31
	EM1: .0, .4	EM2: .0, .4	EM3: .0, .4	EM4: .0, .4
	EC	EC	EC	EC
sum of cell energies	32, 33	34, 35	36, 37	38, 39
energy fractions	40, 41	42, 43	44, 45	46, 47
number of cells	48, 49	50, 51	52, 53	54, 55

Table 2: Figure number guide for cell-info plots.

5 Brief Neural Net description

We use Object Oriented implementation of MLPfit package [9] available in ROOT via TMultiLayerPerceptron class [10]. We use 4.00/04 version of ROOT under t04.04.00 DØ Run II release version. This implementation uses simple feed-forward neural network that consists of input nodes, hidden nodes with sigmoid transfer functions and output nodes. 75% of the input sample is used for training and 25% for testing the neural net. Several learning methods are available. We tried such methods as stochastic minimization, steepest descent, and Broyden, Fletcher, Goldfarb, Shanno (BFGS) method using default parameter settings for each method. In the end we decided to use BFGS for NN studies. The code for the trained NN can be written out as a class in the output source and header files to be used in the analysis.

6 Cross-checking NN with H-Matrix variables

We begin with cross-checking NN performance with that of the "HMatrix + track isolation" EMID using HMatrix variables and track isolation as NN input variables. HMatrix-8 variables are: EM floor energy fractions (4 variables), transverse widths in EM3 (2 variables), $\log(\text{energy})$, $Z(\text{Primary Vertex})/\sigma$. In case of HMatrix-7 only one of the transverse widths ($r\phi$ -width) is used. The $Z(\text{Primary Vertex})/\sigma$ variable is needed because HMatrix training is done separately for each η bin corresponding to a calorimeter tower. For NN training we use the whole η range and therefore do not use this variable. We expect NN to have somewhat better S/B than "HMatrix + track isolation" (cuts defined in Section 1) for the following reasons:

1. HMatrix assumes Gaussian distributions of variables while the variables that are used are not Gaussian-distributed. Neural Net input variables do need to be Gaussian-distributed.
2. NN takes into account the correlations between HMatrix and track isolation while "HMatrix + track isolation" does not.
3. Real data is used for NN training while MC is used for HMatrix training.
4. Trained NN has a knowledge of both the signal and the background while HMatrix is trained on signal only.

The only disadvantage of the NN net in comparison with the HMatrix is that the NN training is done for the whole η range (no binning). We start with "ala HMatrix-7" NN for CC and try three learning methods: stochastic, steepest descent, BFGS. Figure 56 shows training curves and NN output for the three methods. Stochastic minimization training curve is not informative about training status. Steepest descent training has not finished yet with the number of epochs that we tried (1500). In case of BFGS both training sample and test sample errors come to a plateau which suggests that NN training has finished [11]. As soon as the plateau starts no more knowledge is added to the NN with more training epochs: the(?) minimum has already been found and extra epochs only correspond to oscillating around the minimum [11].

The values of NN output are not bound between 0 and 1 as expected. This is a bug in ROOT [12] which has no effect on the NN performance. It only affects the overall scale but not the shape of the signal and background NN output.

We compare the NN performance with the training terminated at the plateau (after 1500 epochs) and just before the plateau (after 800 epochs). Figure 57 shows training curves and NN output for the "ala HMatrix-7" CC NN traind with 1500 and 800 epochs. In EC we train NN with HMatrix-8 variables plus track isolation minus Z(Primary Vertex)/ σ . Figure 58 shows training curves and NN output for the "ala HMatrix-8" EC NN traind with 1500 and 500 epochs.

Anoter concern of NN training is avoiding overtraing. Overtraining means that NN memorizes the whole sample. Such a NN is not capable of classifying an event on a basis of "patterns" and can not be used in the analysis. Figure 59 shows an example of overtrained NN. In this example input sample consists of 50 events. As the number of training epochs increases, the NN memorizes the sample: training sample error goes to zero while test sample error increases.

When the training is performed with the full sample of 10K events, the 7(8)-variable NN does not get overtrained as Figures 57 and 58 show.

To make a comparison, first, we determine the NN output cuts such that provide the same signal efficiency as "HMatrix + track isolation" EMID. For calculating signal efficiency we use "tag/probe" method described in [1] (page 16).

Figures 60 and 61 show diEM mass distributions with tag and probe objects for CC and EC "HMatrix + track isolation" EMID efficiency measurements respectively. EMID Efficiencies are found to be:

$$86.0 \pm 0.6(stat) \pm 2.7(syst)\% \text{ (CC)}$$

$$91.0 \pm 0.6(stat) \pm 3.4(syst)\% \text{ (EC)}$$

Figures 62 and 64 show diEM mass distributions with tag and probe objects for CC and EC NN efficiency measurements respectively (1500 training epochs). The NN output cuts are tuned to provide similar efficiencies to those of "HMatrix + track isolation" EMID cuts. Cutting on NN output at 0.1 (CC) and -0.1 (EC) provides the following signal efficiencies:

$$85.5 \pm 0.6(stat) \pm 1.1(syst)\% \text{ (CC)}$$

$$90.7 \pm 0.5(stat) \pm 1.6(syst)\% \text{ (EC)}$$

NN output for the probe object corresponding to this measurement is shown in Figures 63 and 65 for CC and EC respectively.

Then we repeat the EMID efficiency measurement for the NN trained with a smaller number of epochs (stopped just before the plateau that can be seen on the corresponding NN training plots on Figures 57 and 58). Figures 66 and 68 show diEM mass distributions with tag and probe objects for CC and EC NN efficiency measurements respectively (NN training terminated just before the plateau: 800(CC) and 500(EC) training epochs). The NN output cuts are tuned to provide similar efficiencies to those of "HMatrix + track isolation" EMID cuts. Cutting on NN output at 0.1 (CC) and -0.1 (EC) provides the following signal efficiencies:

$$85.6 \pm 0.6(stat) \pm 0.7(syst)\% \text{ (CC)}$$

$$92.1 \pm 0.5(stat) \pm 1.6(syst)\% \text{ (EC)}$$

NN output for the probe object corresponding to this measurement is shown in Figures 67 and 69 for CC and EC respectively.

The values of the NN cuts were tuned so that the efficiencies are within one standard deviation with those for "HMatrix + track isolation" EMID. The fake rate is not expected to be sensitive to the variation of the cuts within a standard deviation window since the background NN output is flat around (-0.1,0.1) region (Figures 57 and 58)

Having established the NN output cuts of similar efficiencies we then estimate the fake rate. The method of fake rate estimation is described in [1] (page 6).

Figure 70 shows the fake rates for "HMatrix + track isolation" and NN for 1500 epochs (on the training plateau). As a cross-check we plot NN output for loose em objects (fake rate denominator) and photons (fake rate numerator) to make sure that these distributions are consistent with NN output for background training sample. Figure 71 shows loose em and photon NN output for the di-EM sample while Figure 72 shows same distribution for the SingleEM sample.

We also calculate the fake rate (for NN training terminated just before the plateau: 800 (CC) and 500 (EC) training epochs). Figure 73 shows the fake rates for "HMatrix + track isolation" and NN. Figure 74 shows loose em and photon NN output for the di-EM sample while Figure 75 shows same distribution for the SingleEM sample.

7 Current conclusions, questions, and plans

1. Cell info profiles (number of cells and energy fractions) provide noticeable discrimination between signal and background. Need to find optimal set of variables and check if it gives extra discrimination power (in addition to calorimeter cluster variables such as widths and isolation).
2. Use BFGS method for NN training.
3. As soon as NN training plateau is reached NN performance is not sensitive to the choice of the number of training epochs.
4. in CC for comparable signal efficiencies NN gives slightly higher fake rate in the middle of CC and slightly lower fake rates towards the edges of the CC than "HMatrix + Track Isolation" EMID. This could be related to η -binning on/off for "HMatrix + Track Isolation" / NN.
5. in EC NN fake rate is systematically higher. For the 2EM sample the difference is huge. Why ?
6. Why NN based on HMatrix variables and track isolation does not perform better S/B-wise than HMatrix + track isolation ?
 - Does NN get into a local minimum ? Can check this by using the same set of variables with likelihood method.
 - Is the number of variable too high so that NN functionality is spoiled while HMatrix does not care ? Can try to reduce the number of variables
 - Is the training method / choice of training parameters not optimal?
 - Is it that the data just does not contain more of S/B-discrimnationg info than "HMatrix + Track Isolation" EMID relies upon ? (most pessimistic possibility)

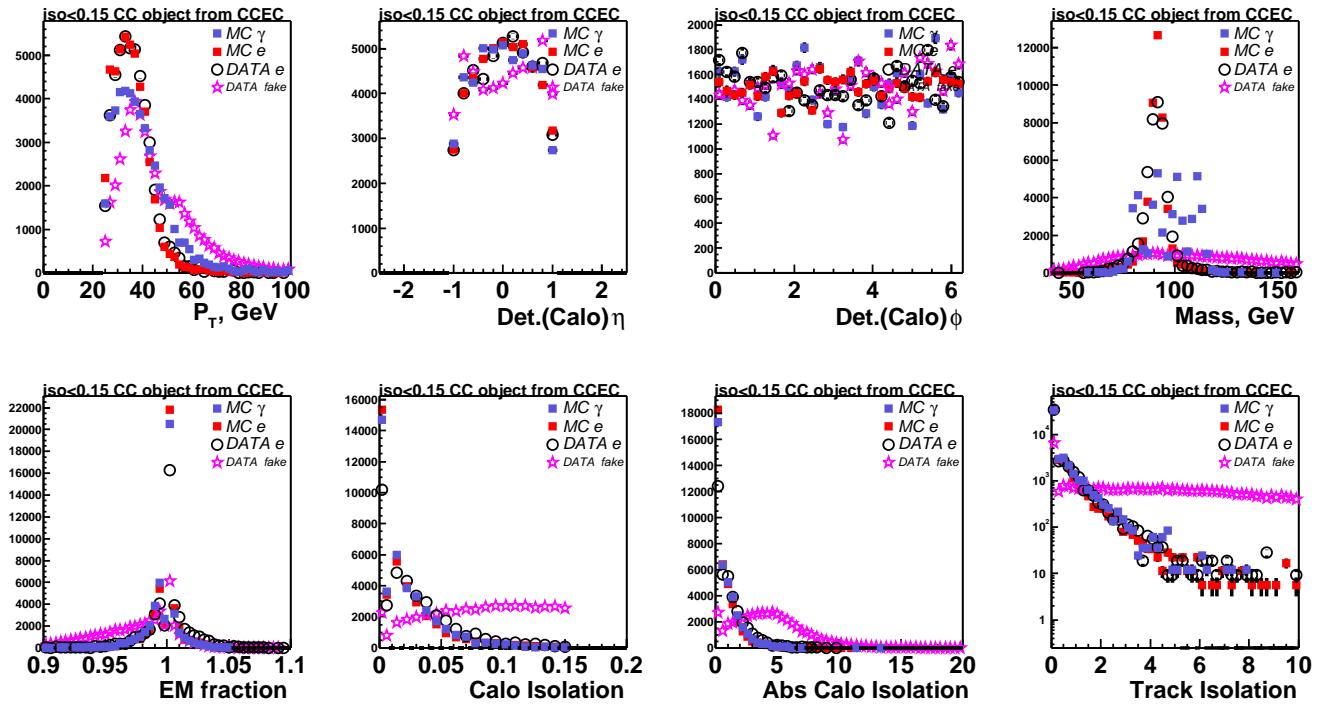


Figure 1: Kinematic and basic EMID variables for isolated ($\text{iso} < 0.15$) CC objects (CCEC)

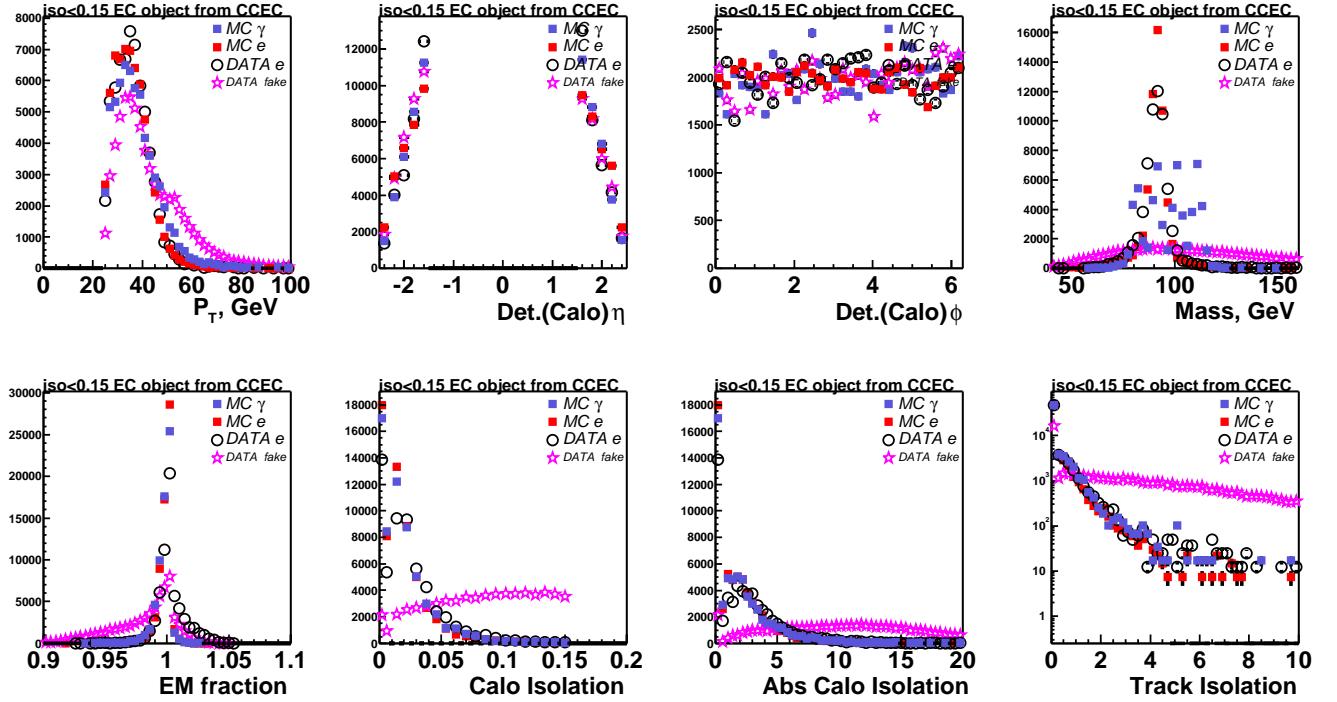


Figure 2: Kinematic and basic EMID variables for isolated ($\text{iso} < 0.15$) EC objects (CCEC)

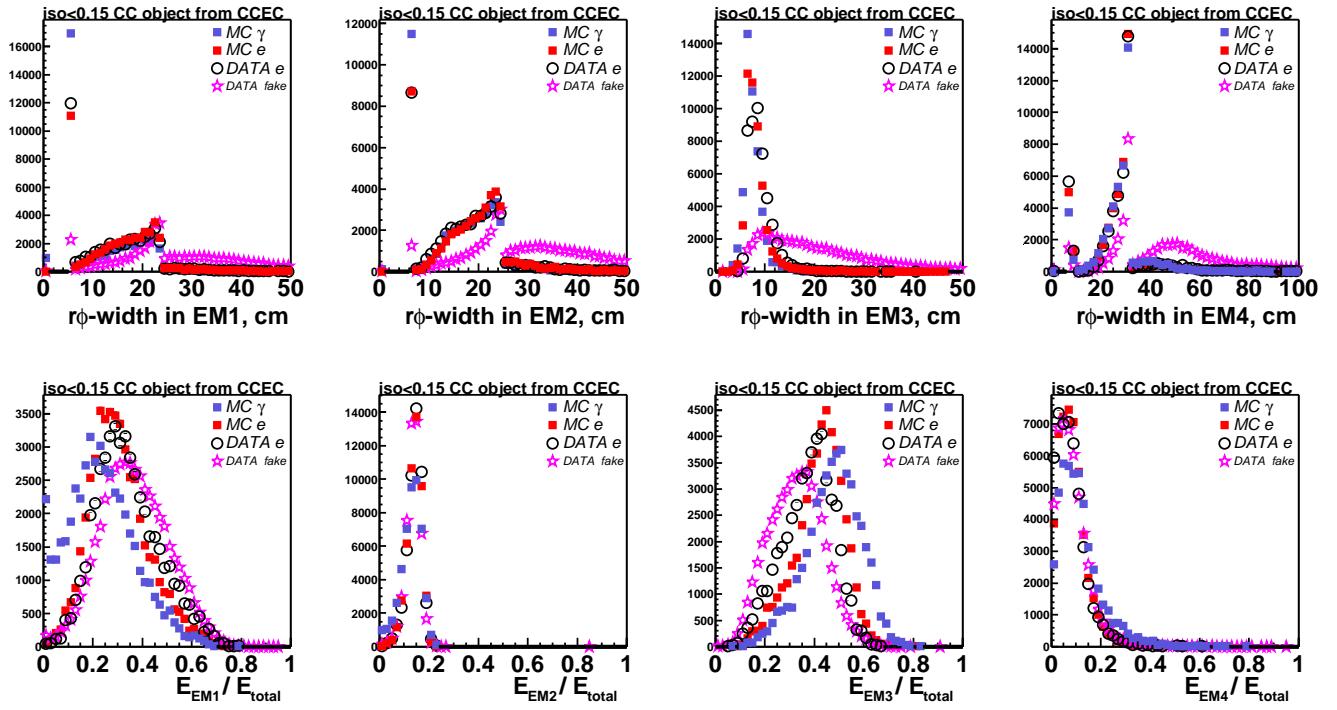


Figure 3: floor $r\phi$ -widths, energy fractions for isolated ($\text{iso} < 0.15$) CC objects (CCEC)

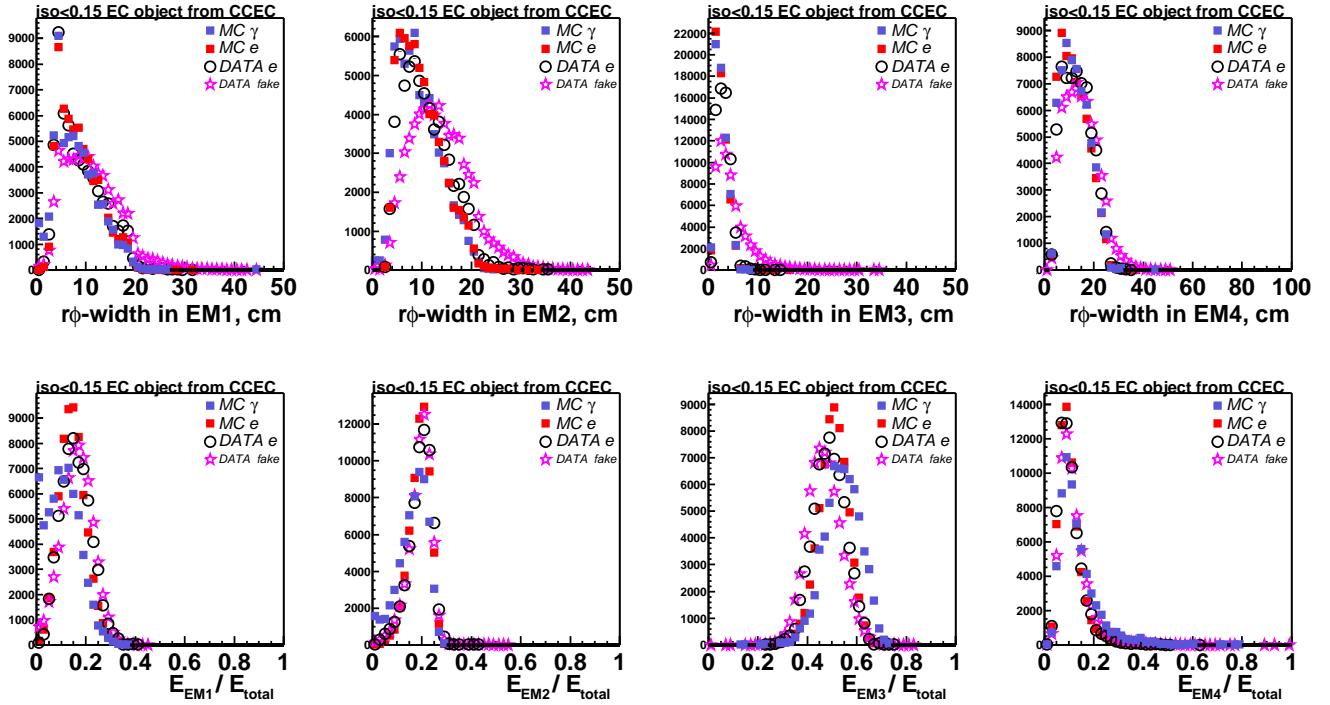


Figure 4: floor $r\phi$ -widths, energy fractions for isolated ($\text{iso} < 0.15$) EC objects (CCEC)

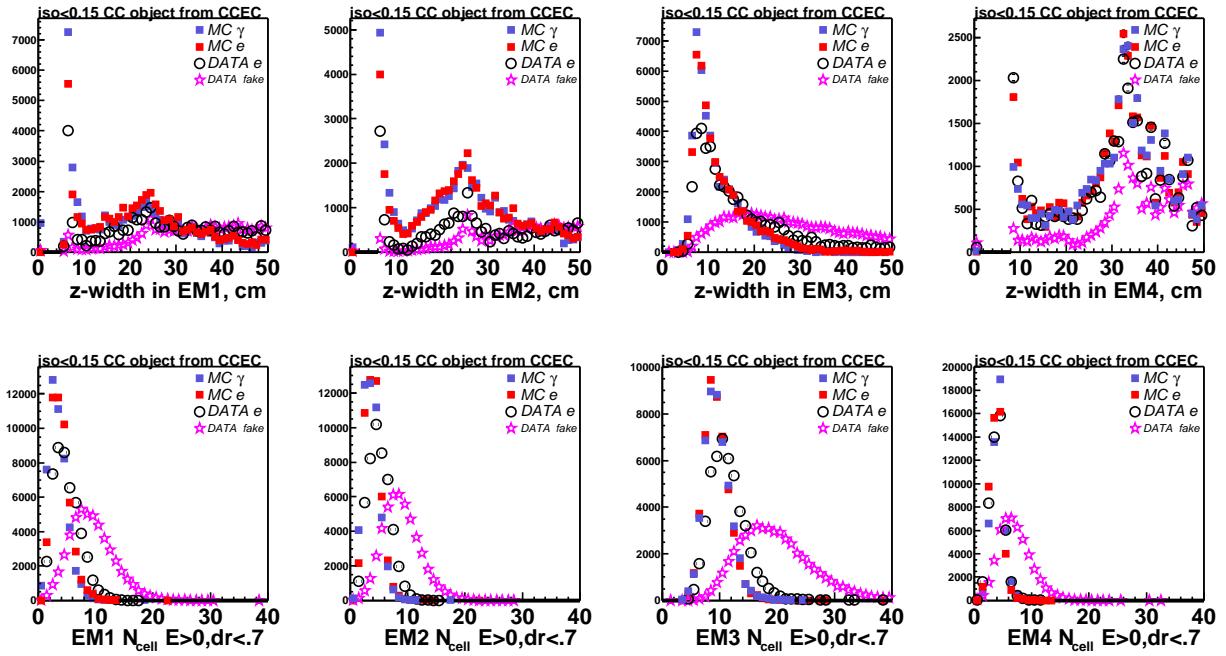


Figure 5: floor z-widths, number of cells in the floor (no cell energy threshold at the analysis stage, $dR(\text{cell}, \text{object}) < 0.7$) for isolated ($\text{iso} < 0.15$) CC objects (CCEC)

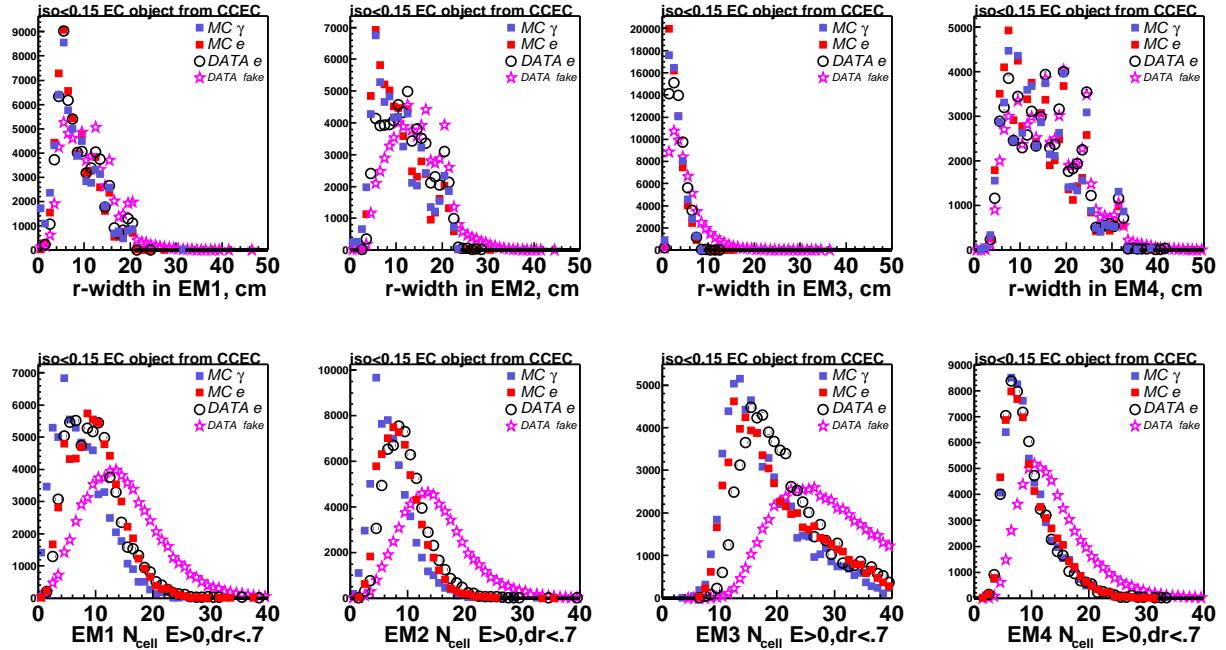


Figure 6: floor r-widths, number of cells in the floor (no cell energy threshold at the analysis stage, $dR(\text{cell}, \text{object}) < 0.7$) for isolated ($\text{iso} < 0.15$) EC objects (CCEC)

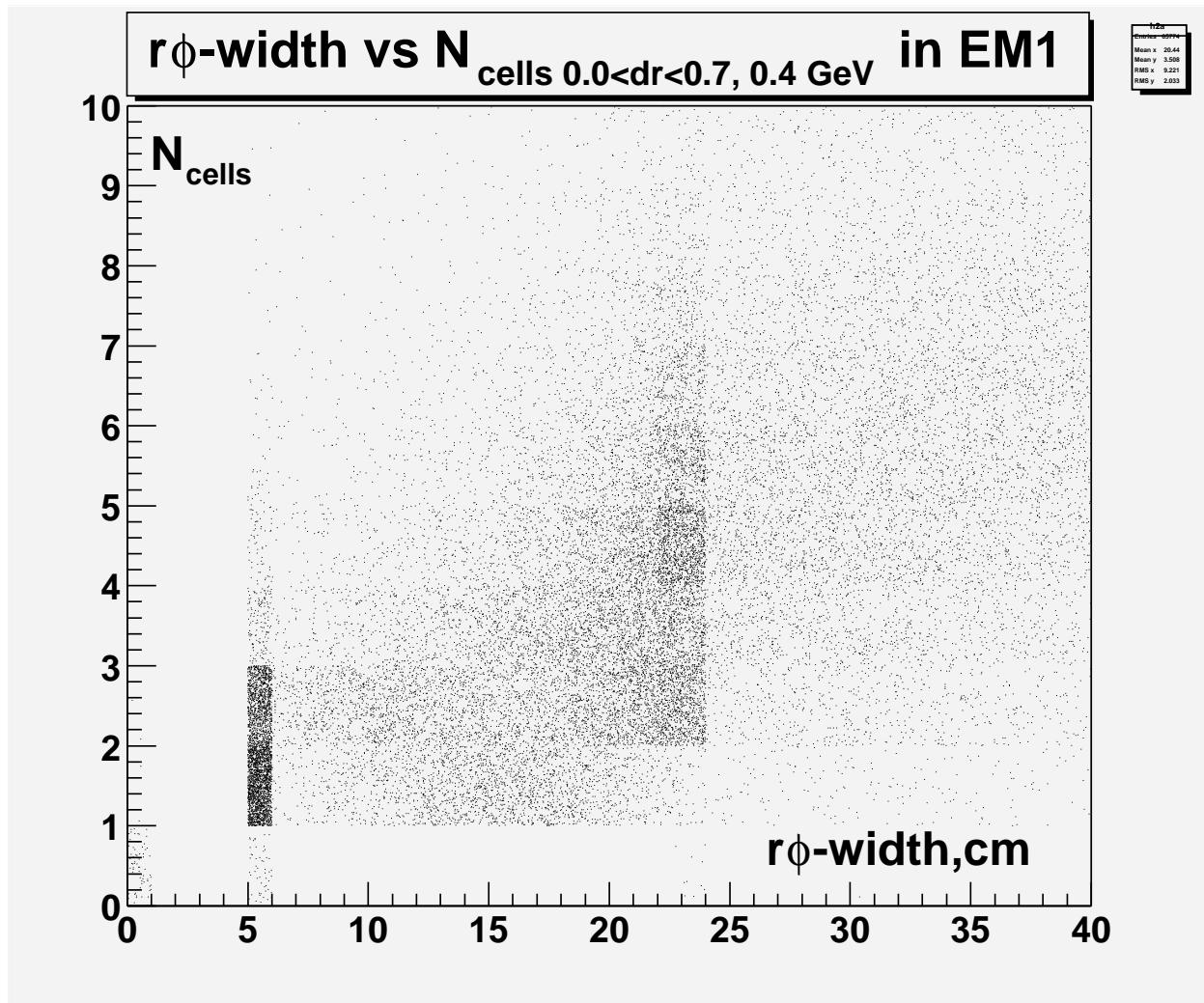


Figure 7: Number of EM1 cells inside dR<0.7 vs. $r\phi$ -width.

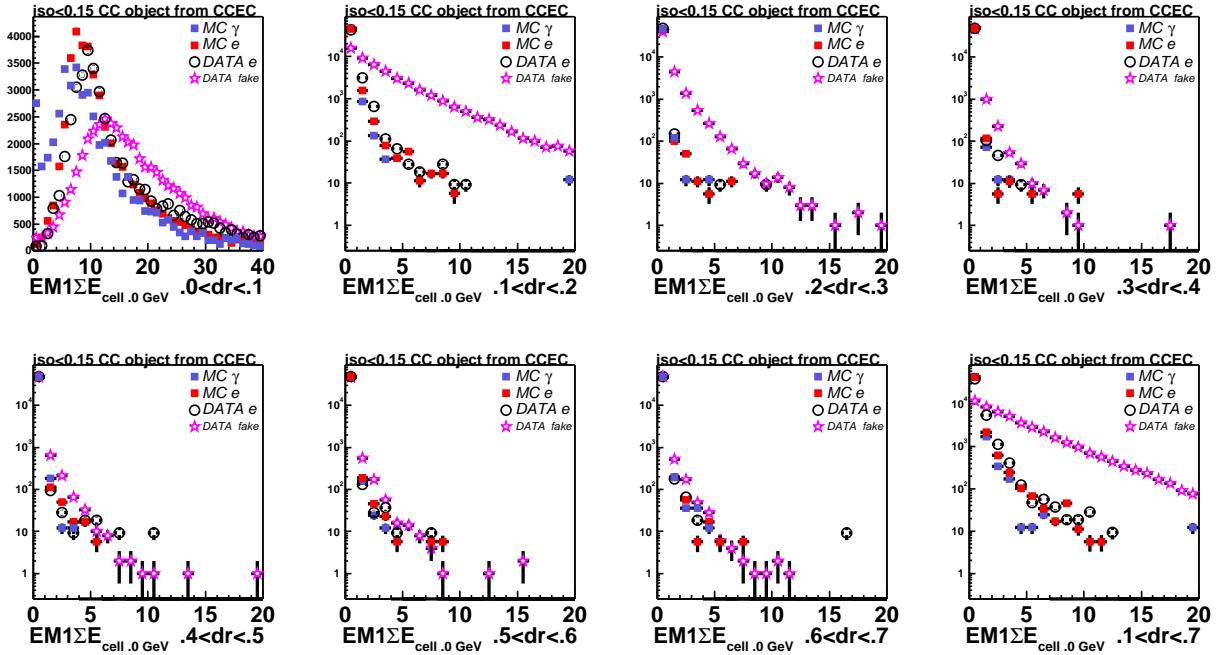


Figure 8: CC EM1: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

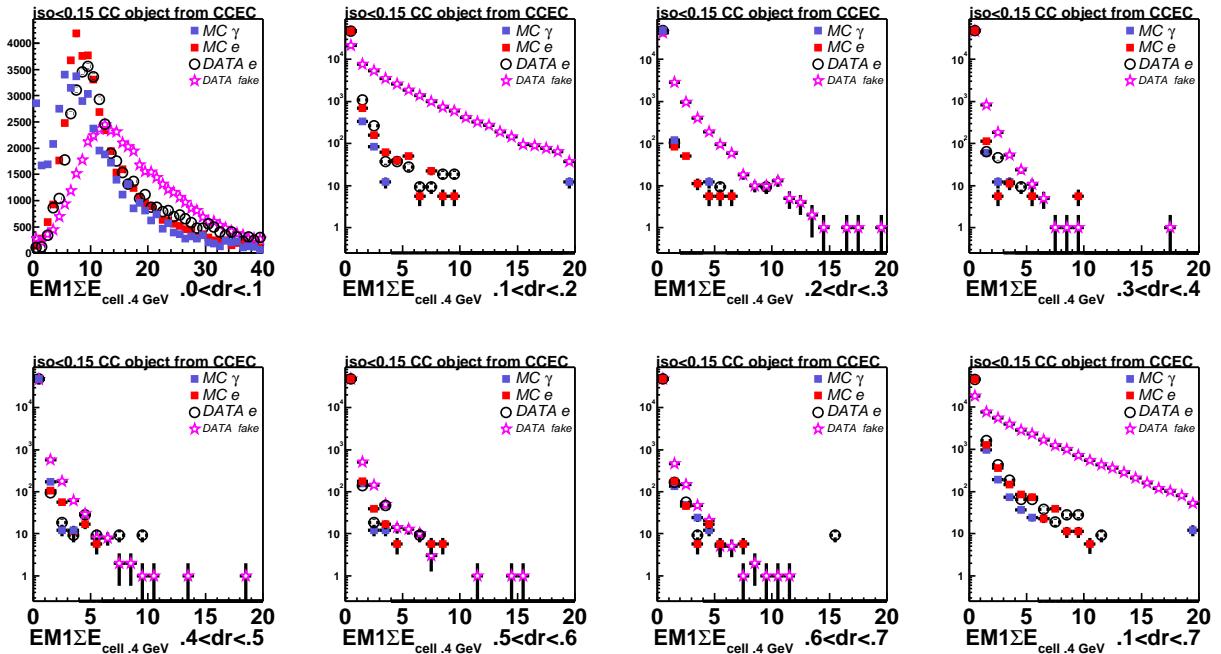


Figure 9: CC EM1: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

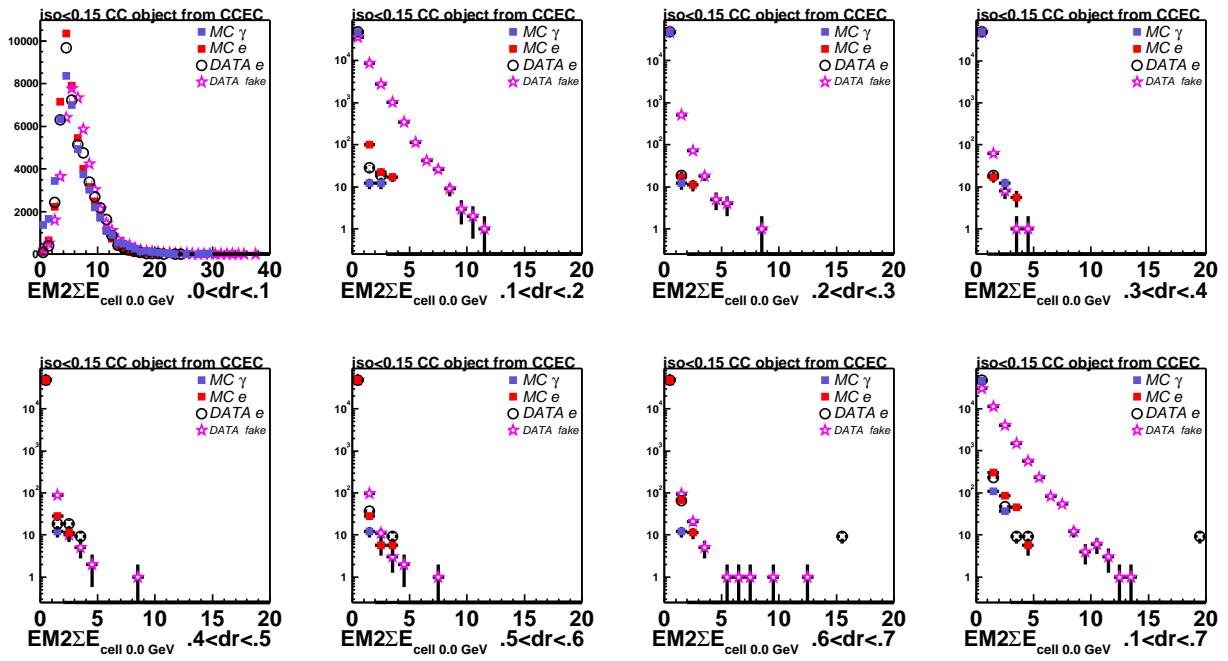


Figure 10: CC EM2: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

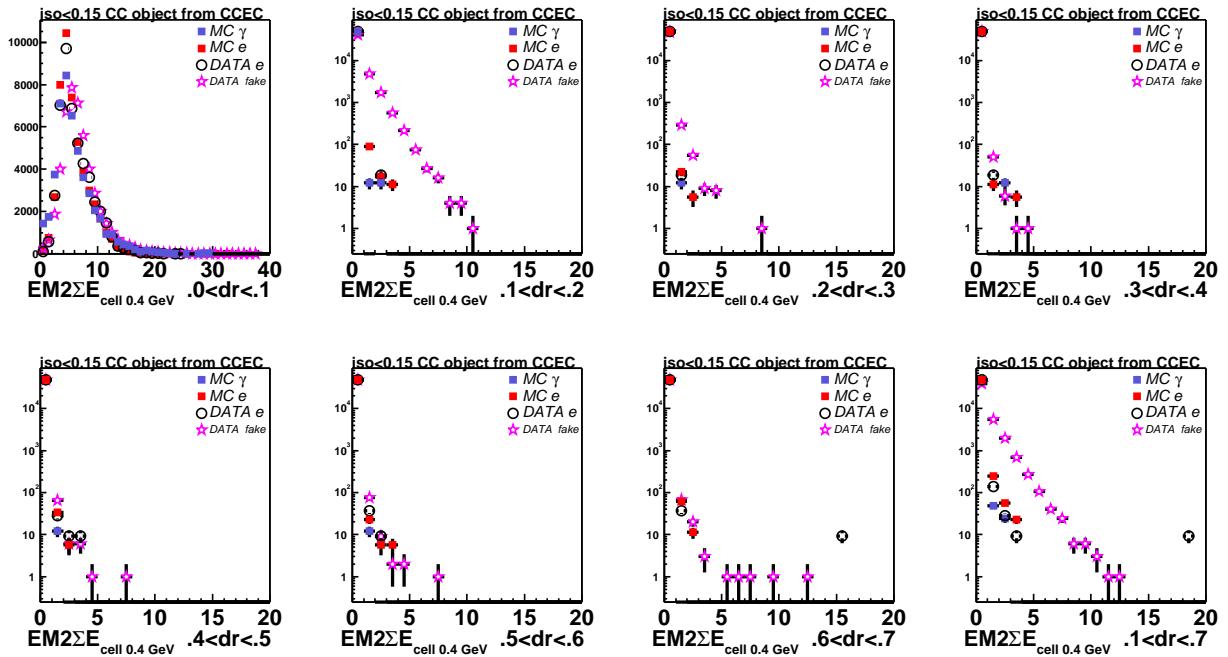


Figure 11: CC EM2: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

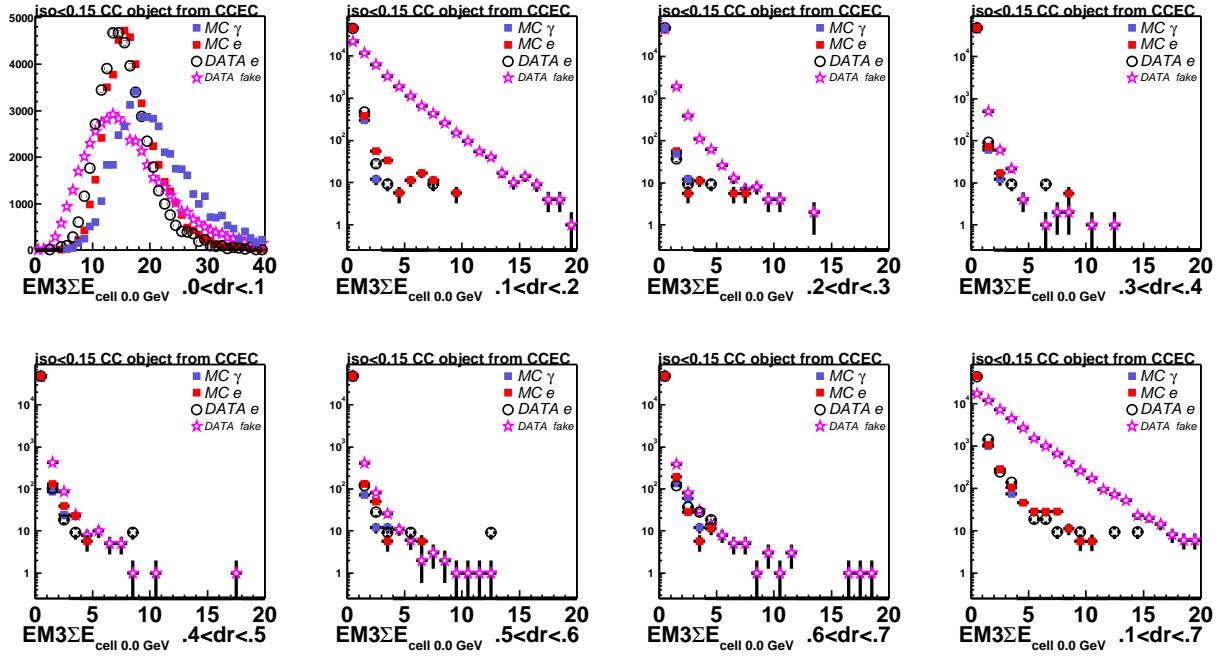


Figure 12: CC EM3: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

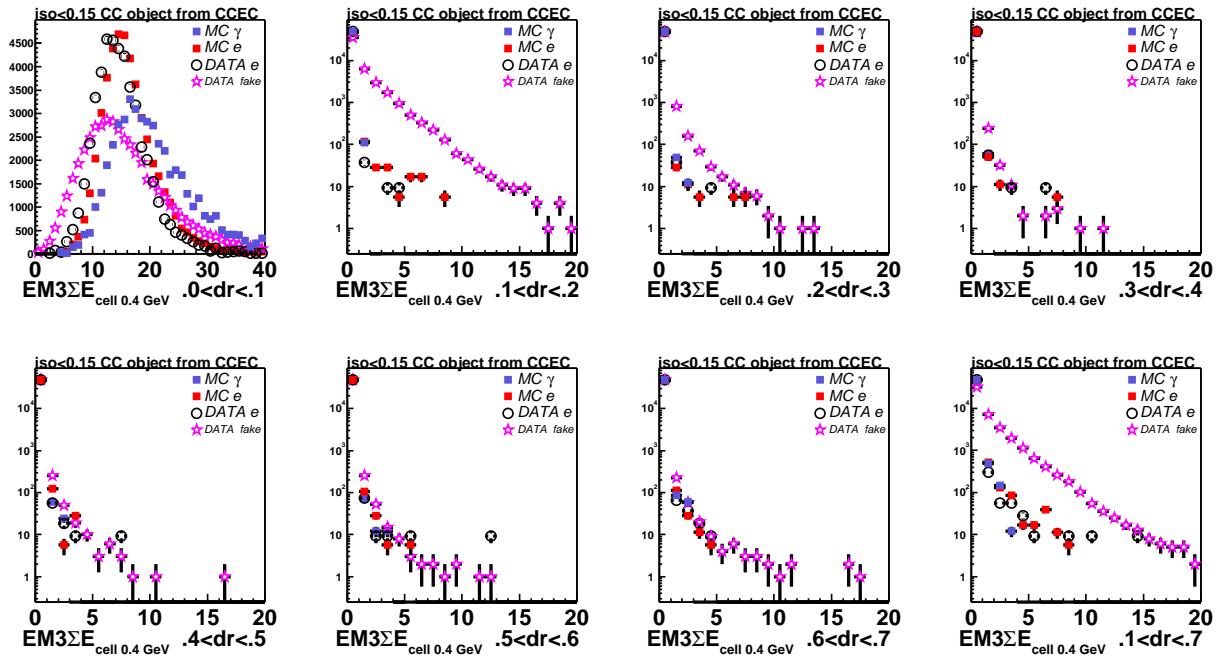


Figure 13: CC EM3: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

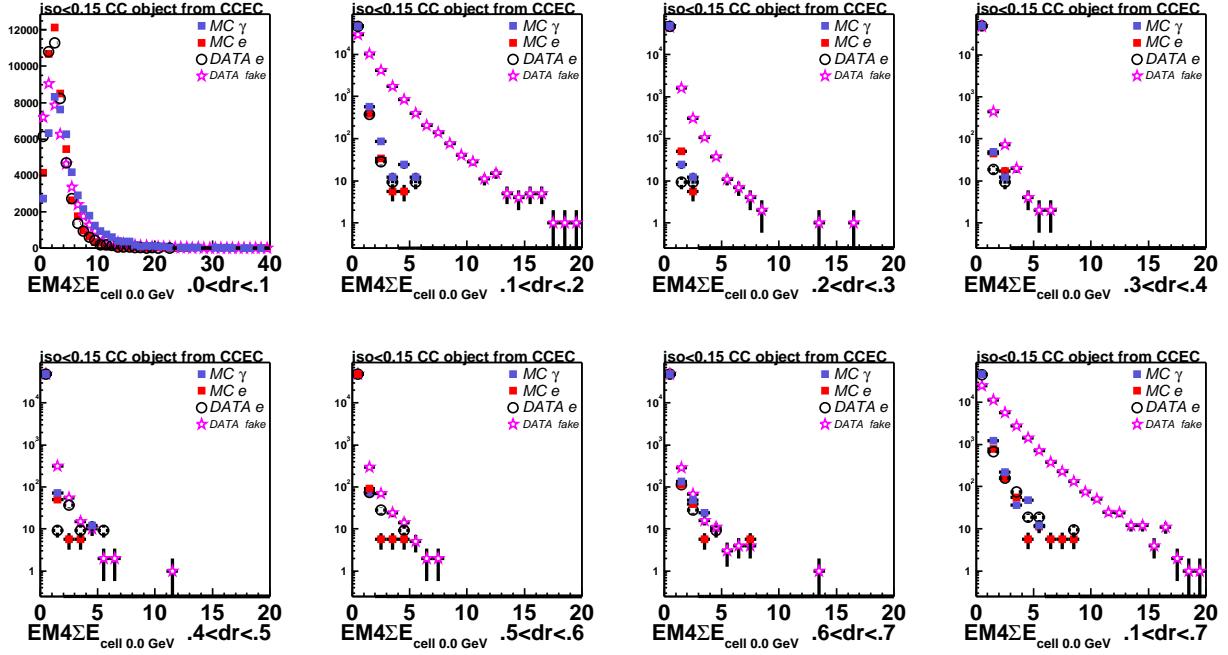


Figure 14: CC EM4: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

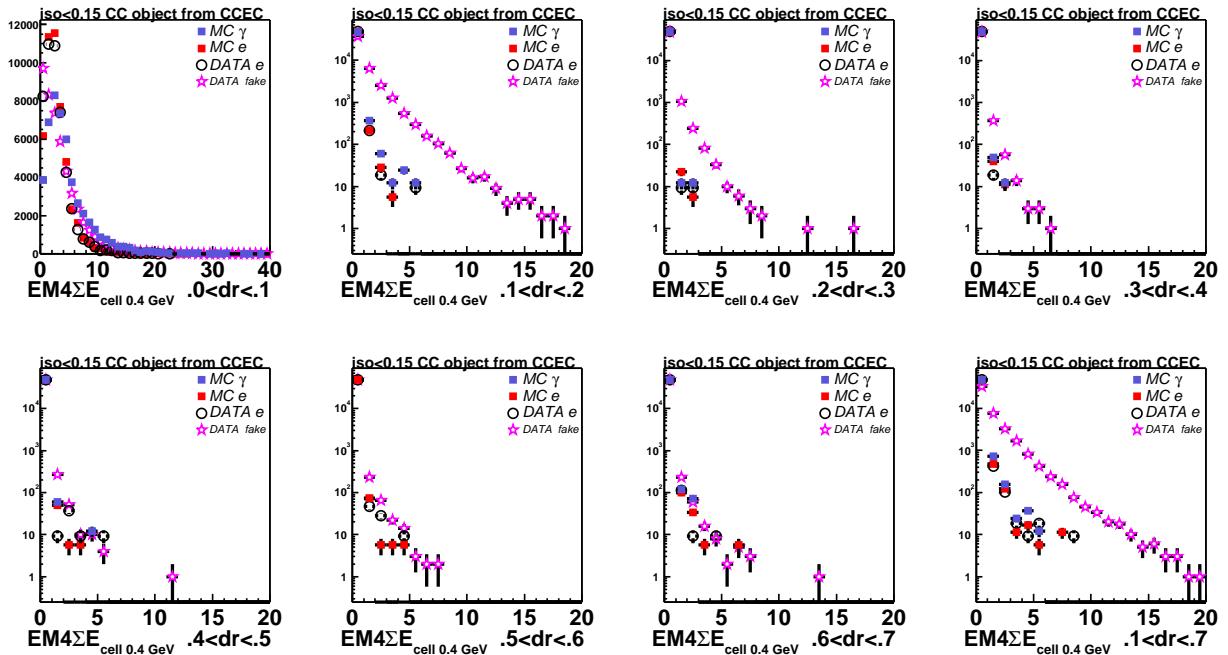


Figure 15: CC EM4: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

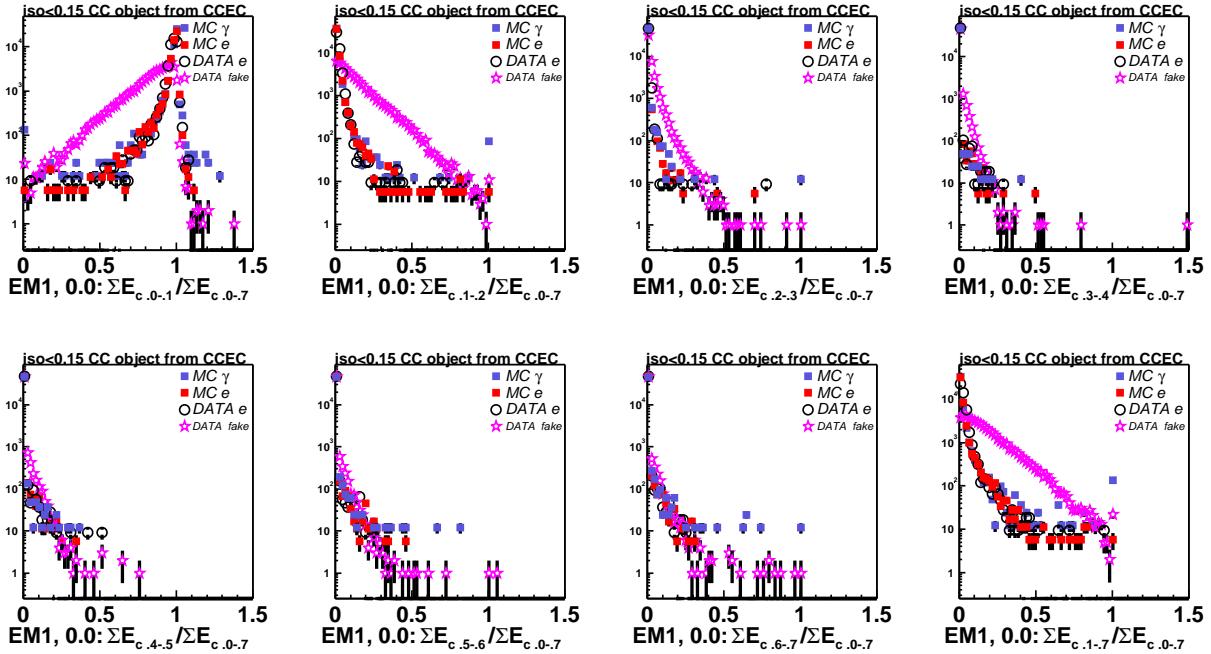


Figure 16: CC EM1: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

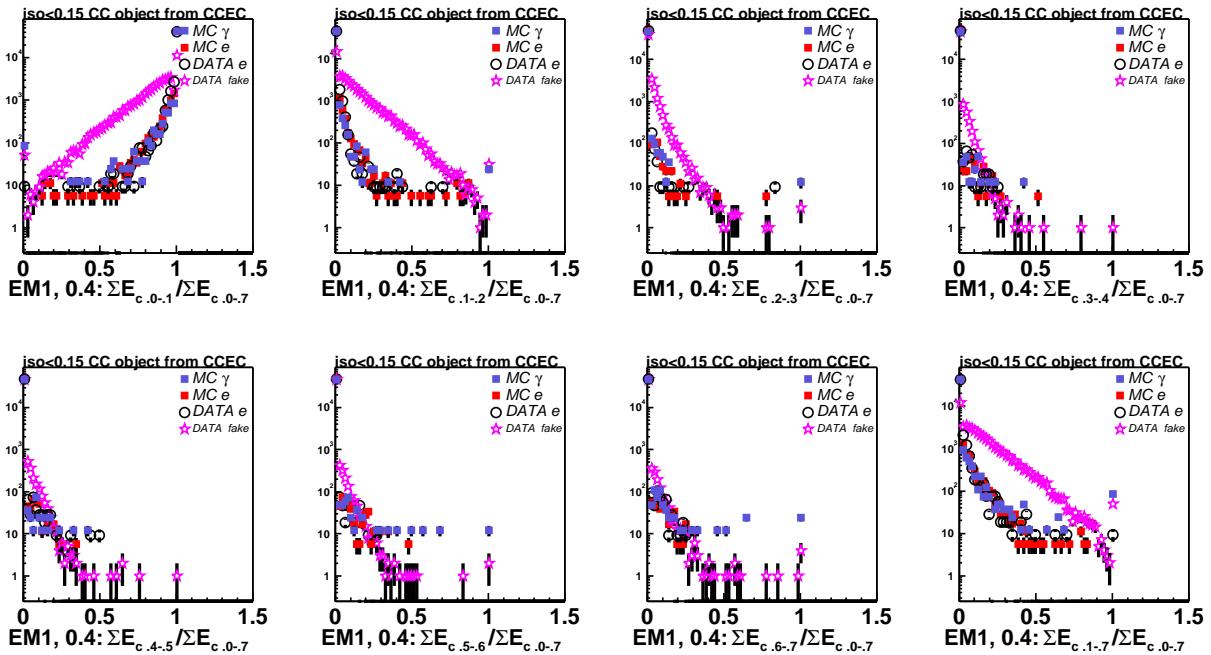


Figure 17: CC EM1: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

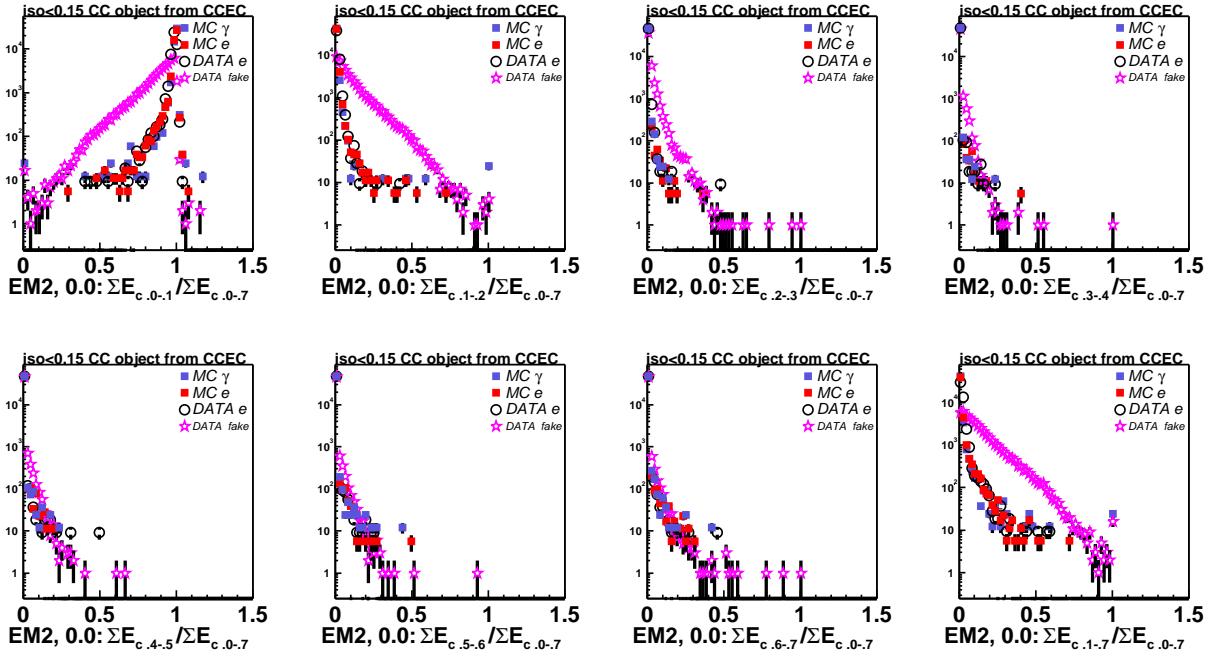


Figure 18: CC EM2: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

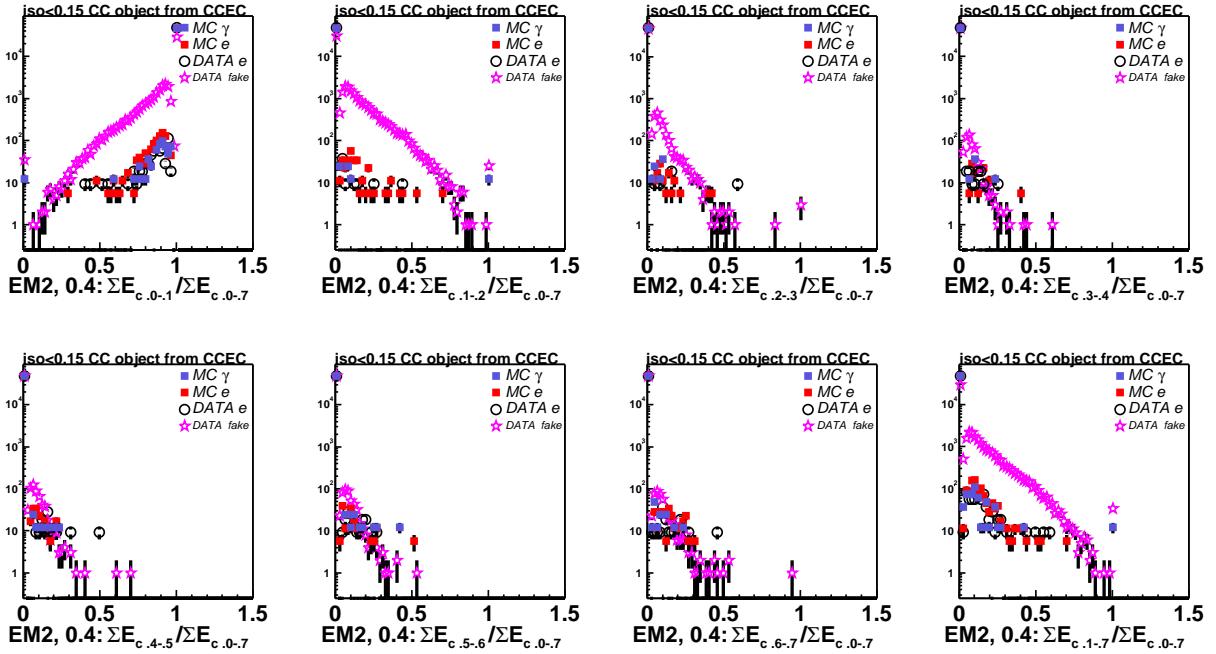


Figure 19: CC EM2: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

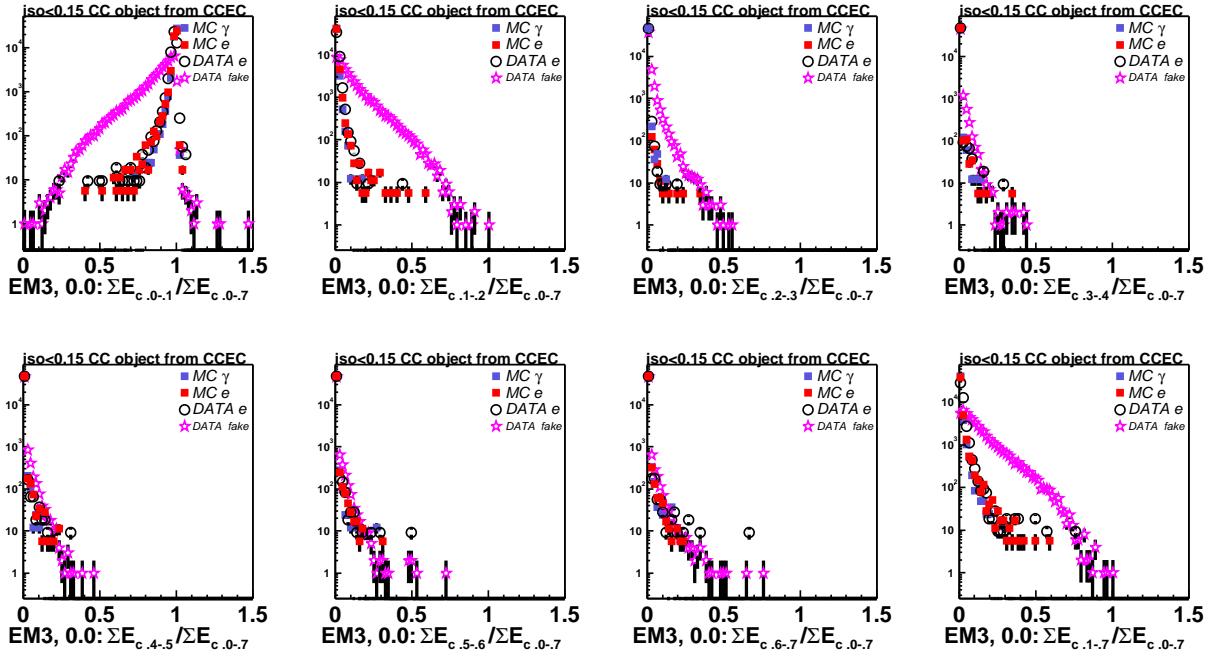


Figure 20: CC EM3: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

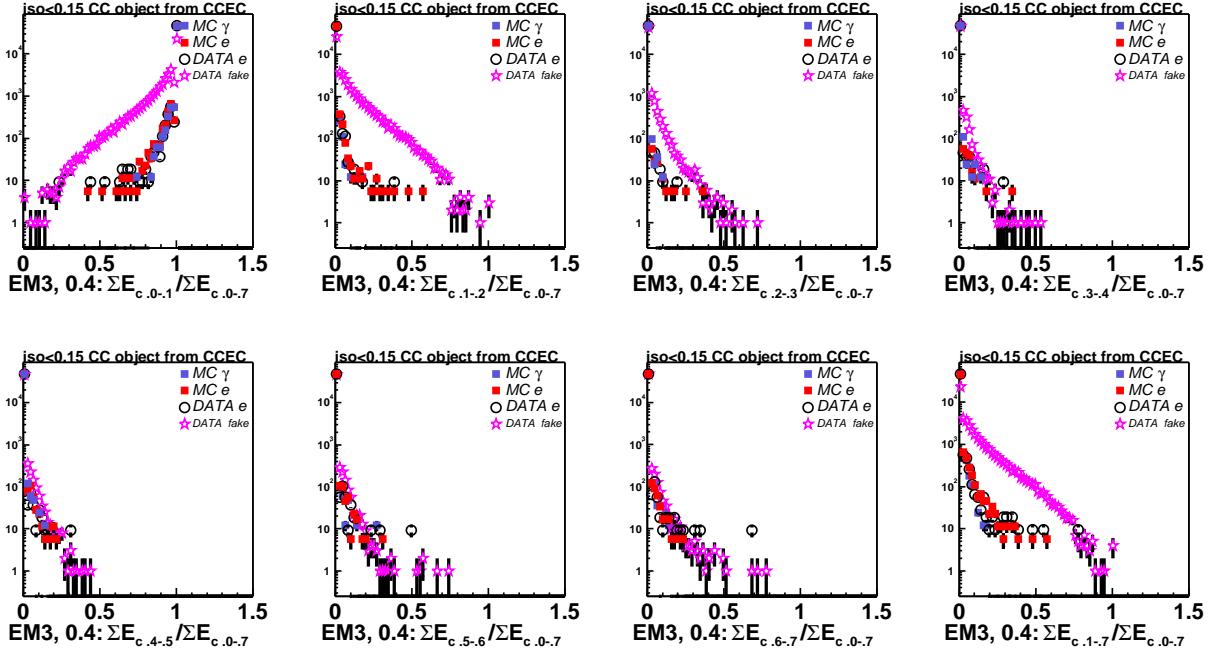


Figure 21: CC EM3: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

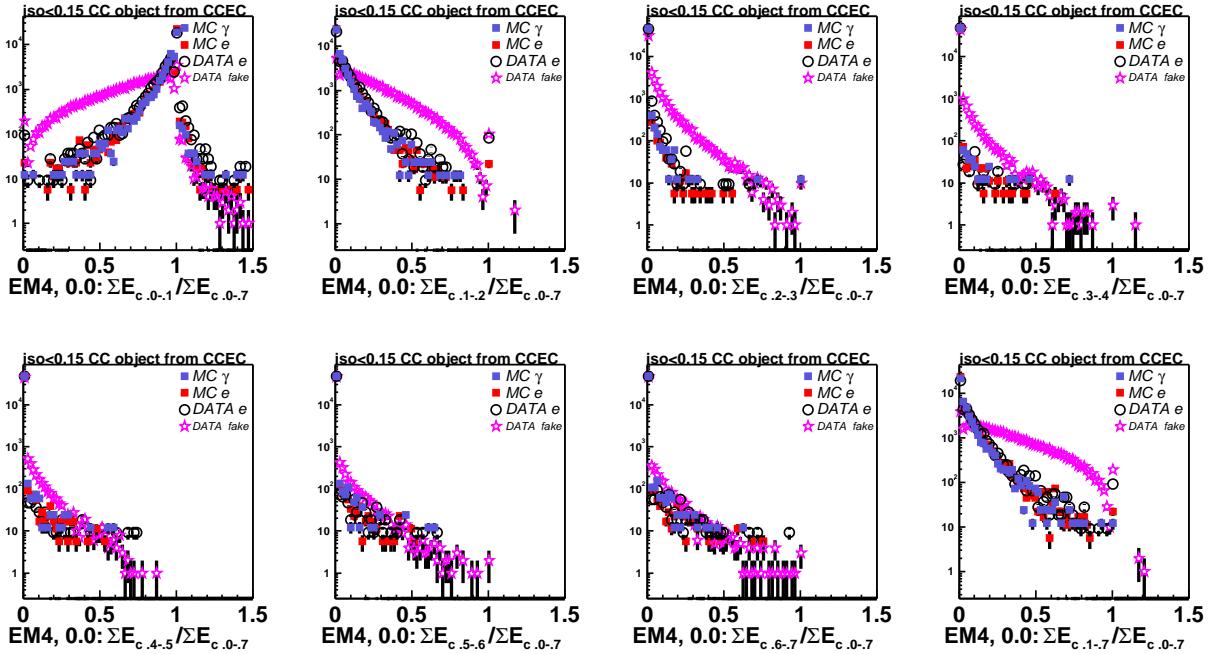


Figure 22: CC EM4: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

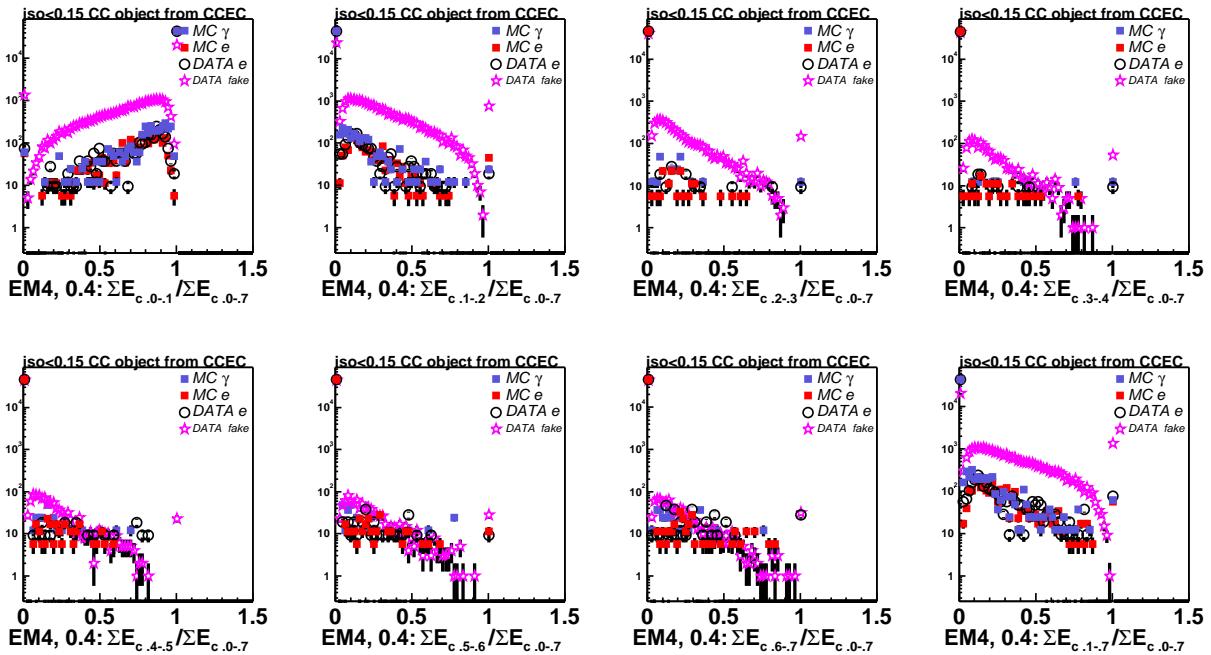


Figure 23: CC EM4: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

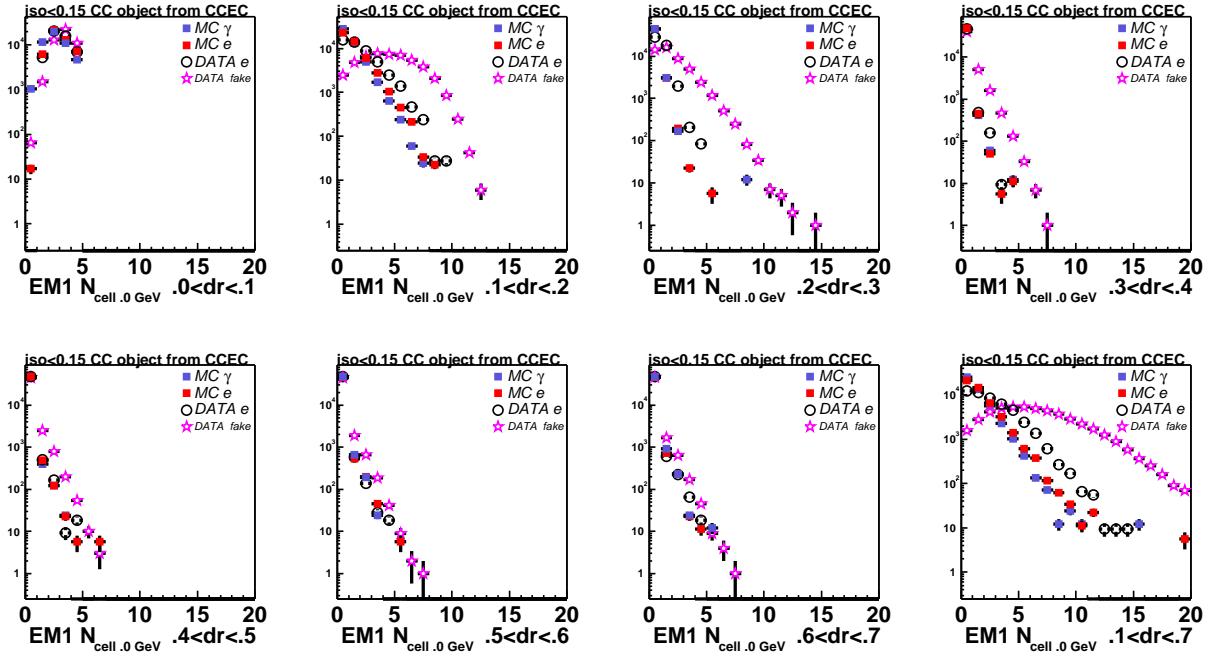


Figure 24: CC EM1: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

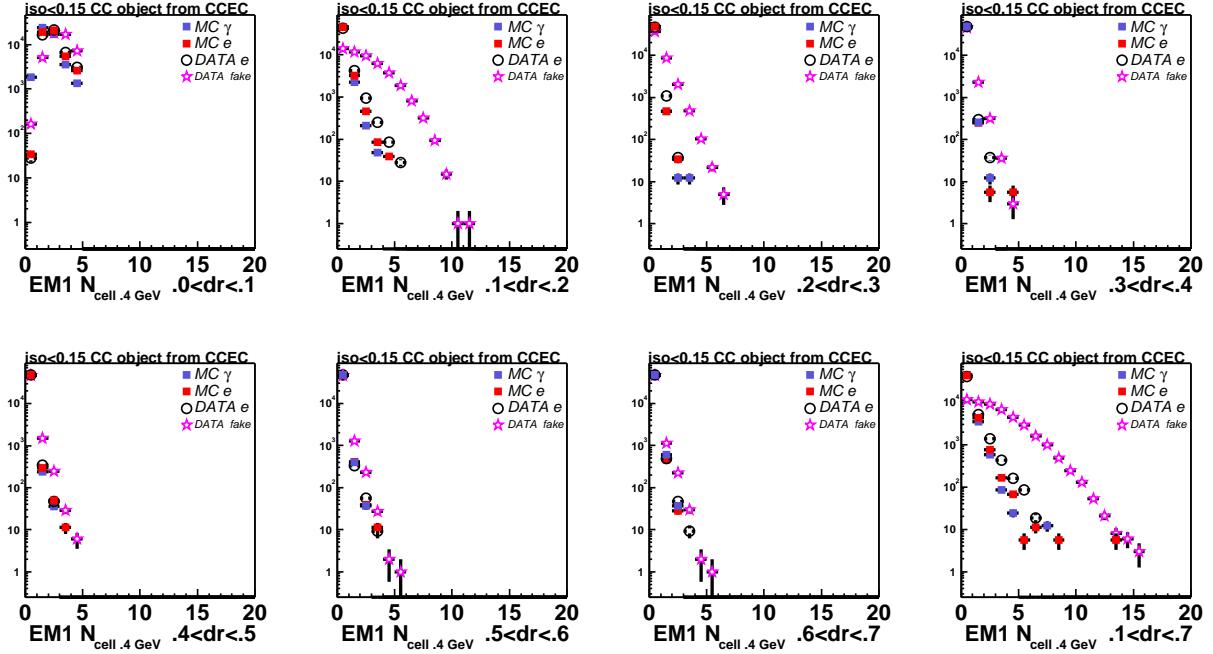


Figure 25: CC EM1: number of cells in dR rings, (cell energy threshold is 400 MeV).

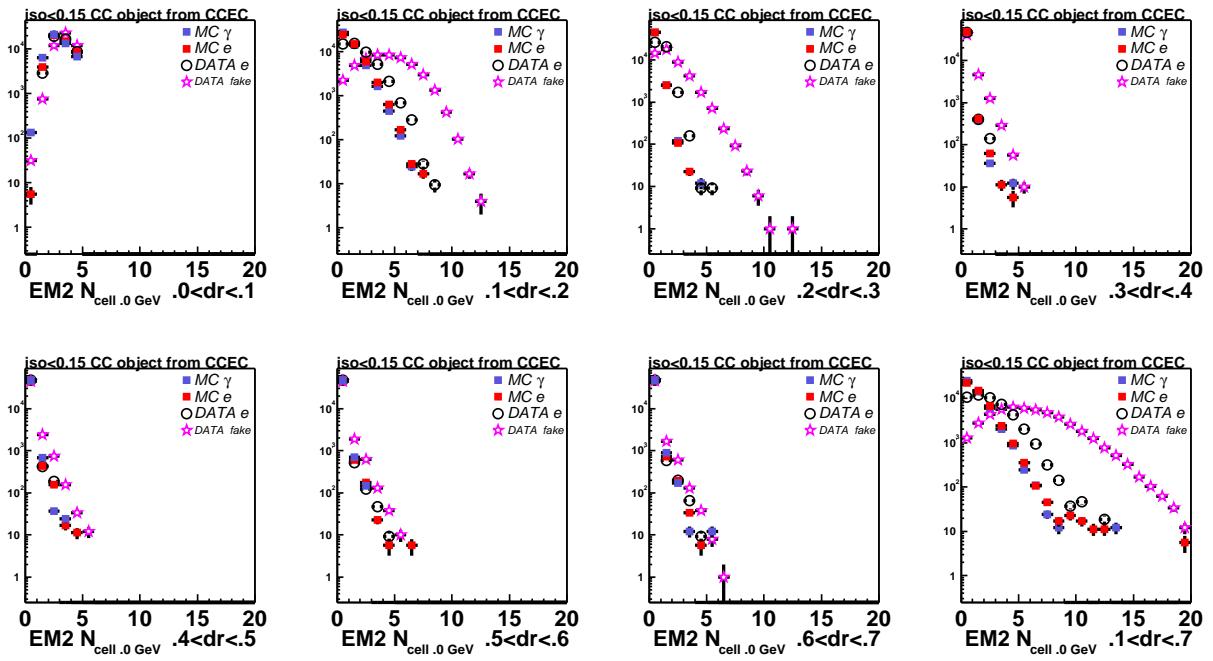


Figure 26: CC EM2: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

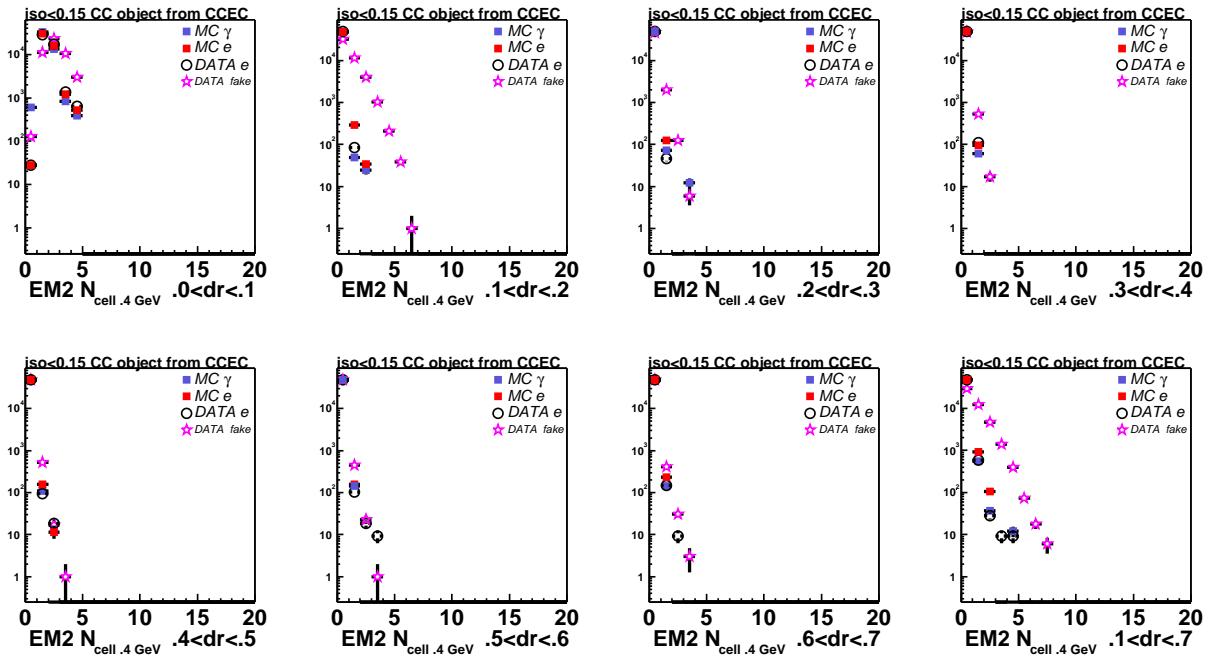


Figure 27: CC EM2: number of cells in dR rings, (cell energy threshold is 400 MeV).

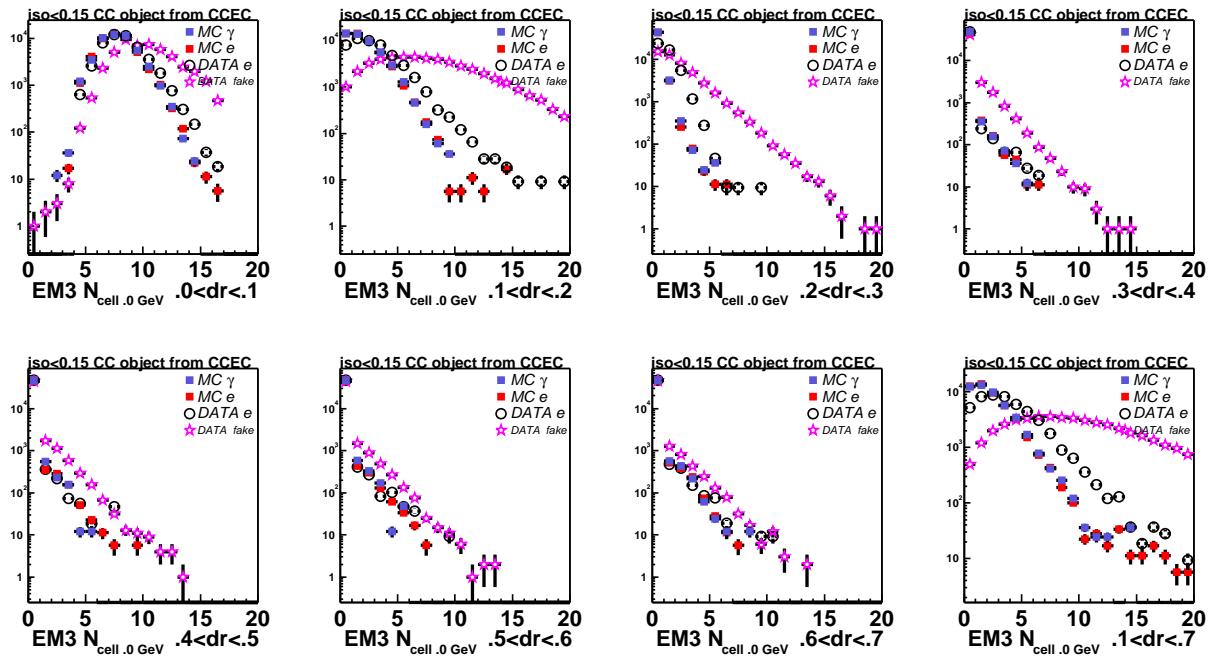


Figure 28: CC EM3: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

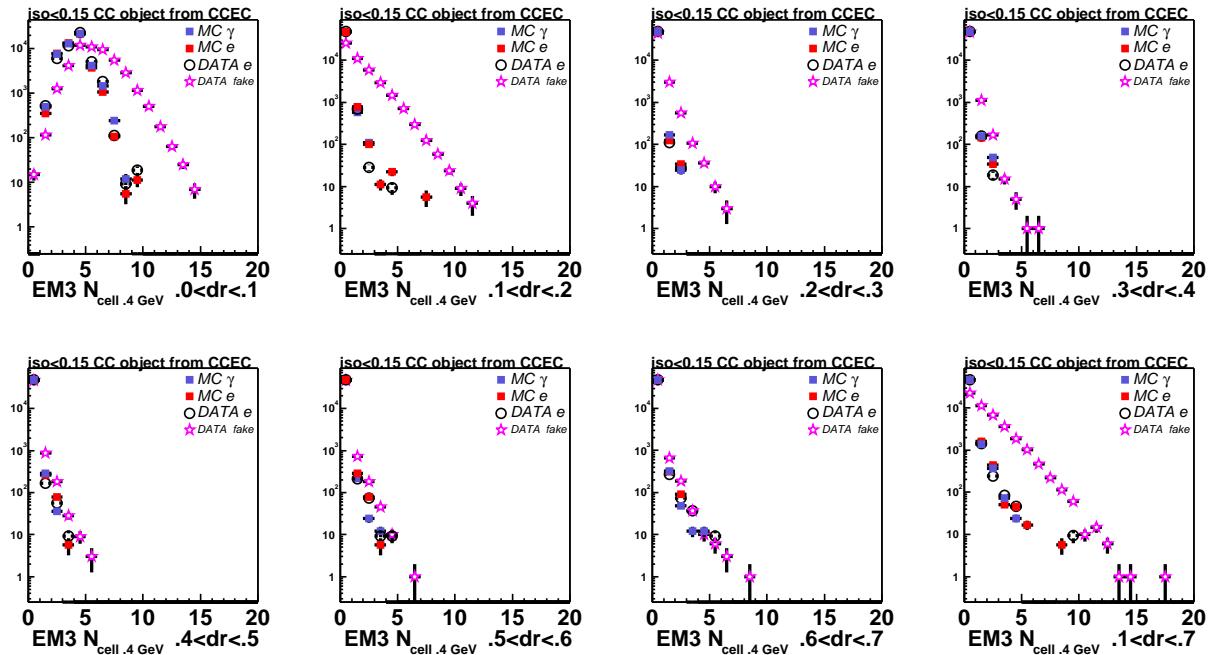


Figure 29: CC EM3: number of cells in dR rings, (cell energy threshold is 400 MeV).

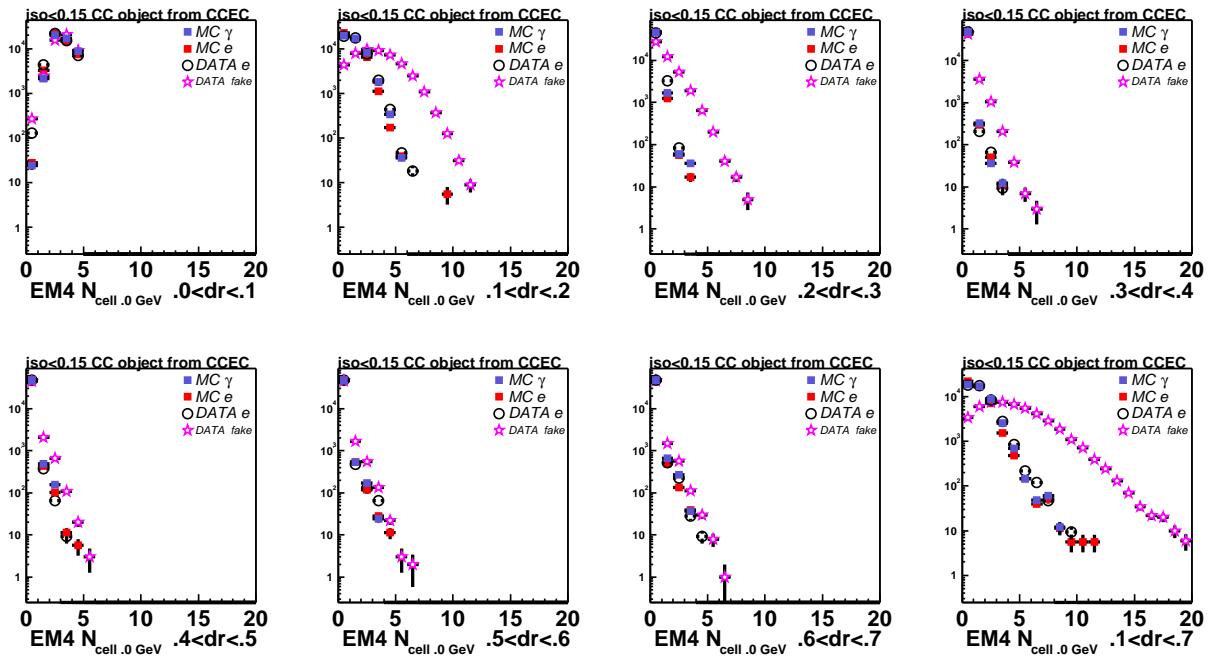


Figure 30: CC EM4: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

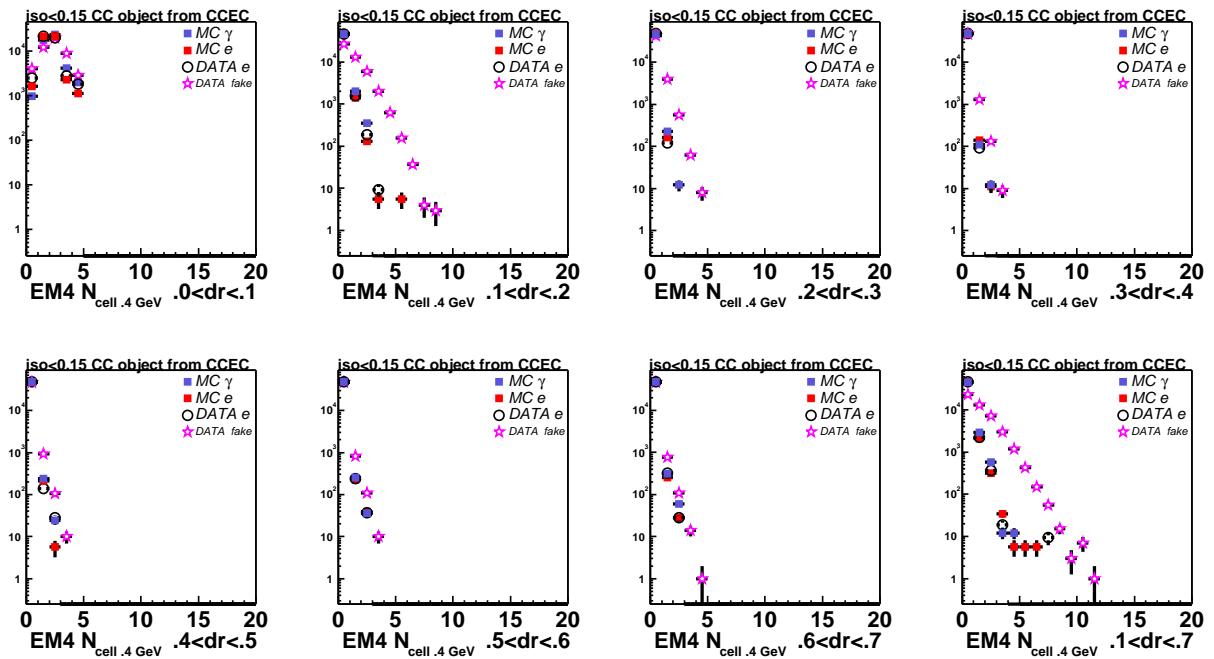


Figure 31: CC EM4: number of cells in dR rings, (cell energy threshold is 400 MeV).

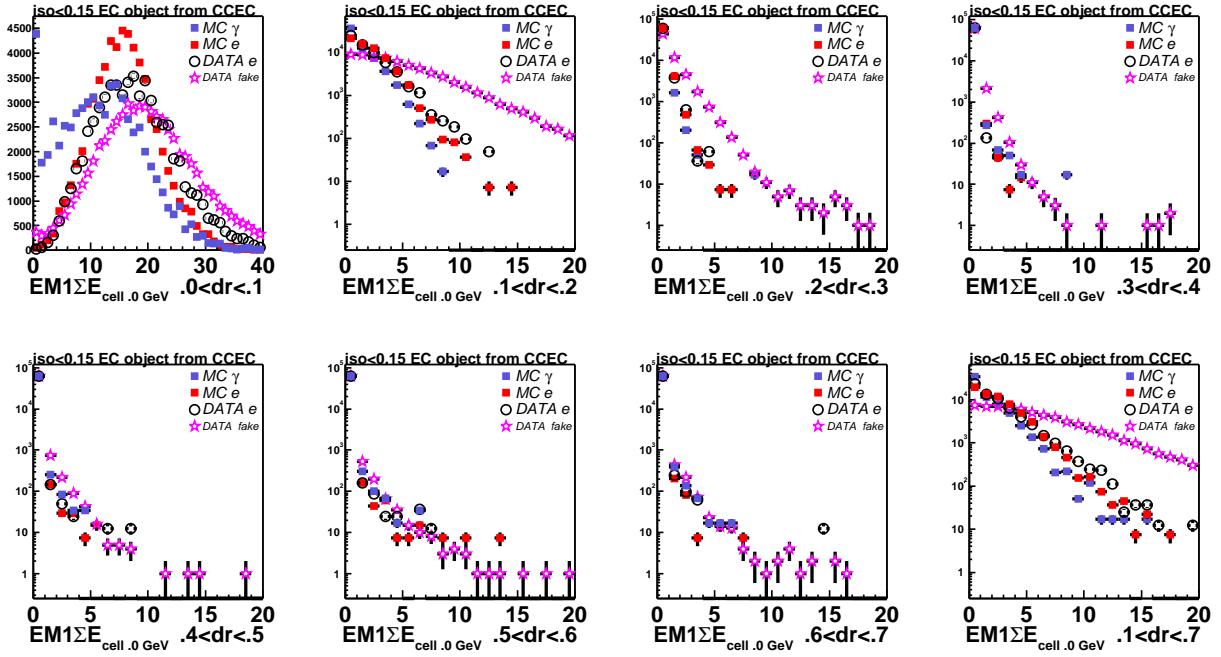


Figure 32: EC EM1: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

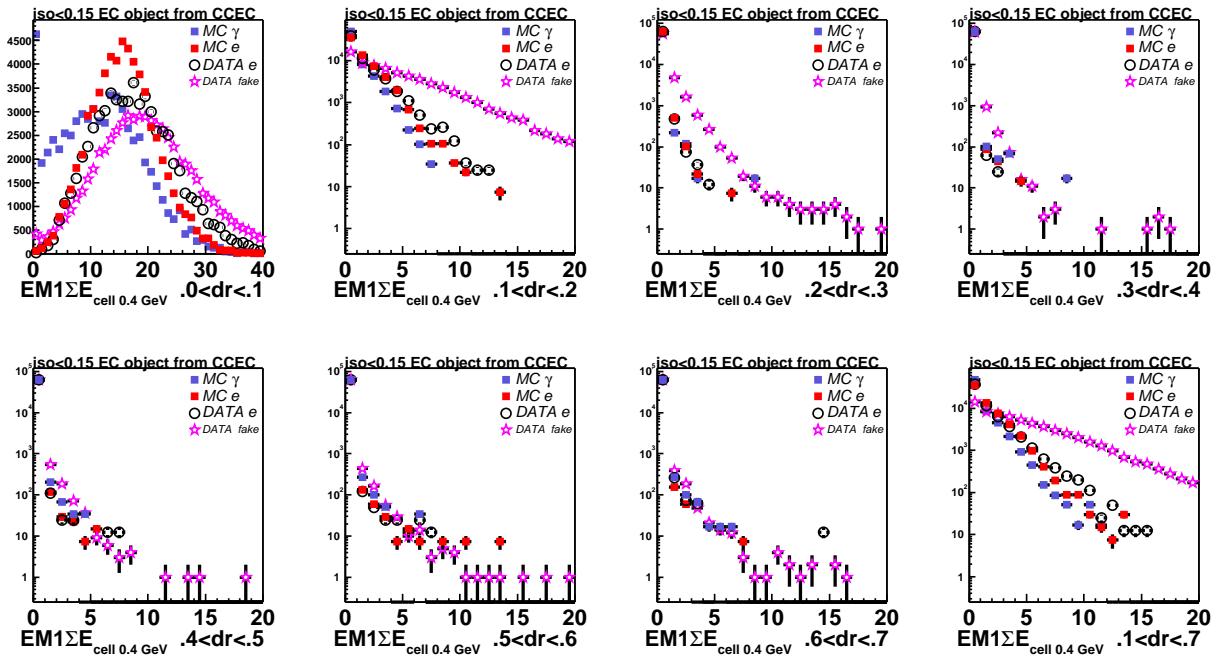


Figure 33: EC EM1: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

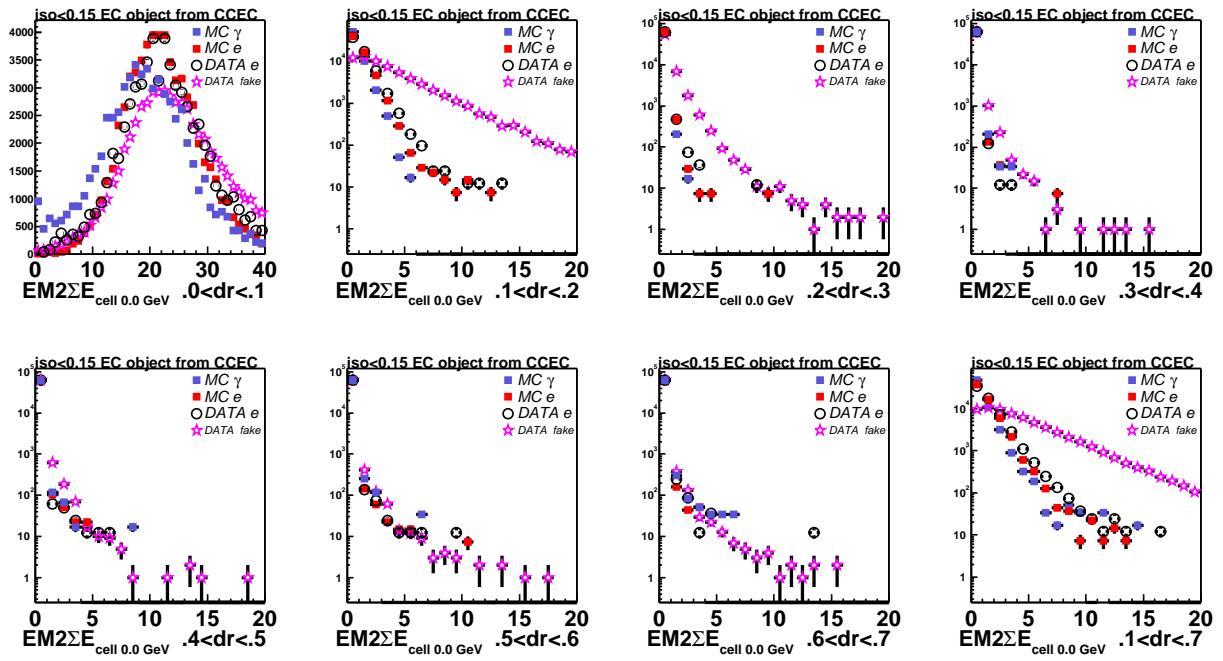


Figure 34: EC EM2: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

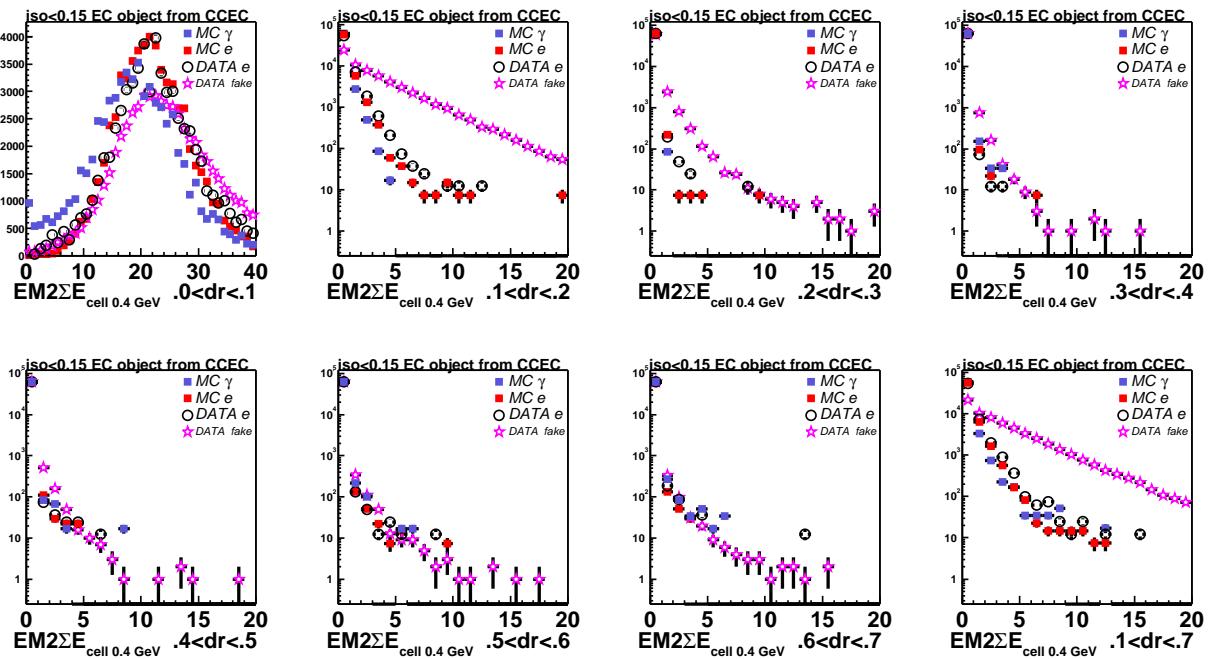


Figure 35: EC EM2: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

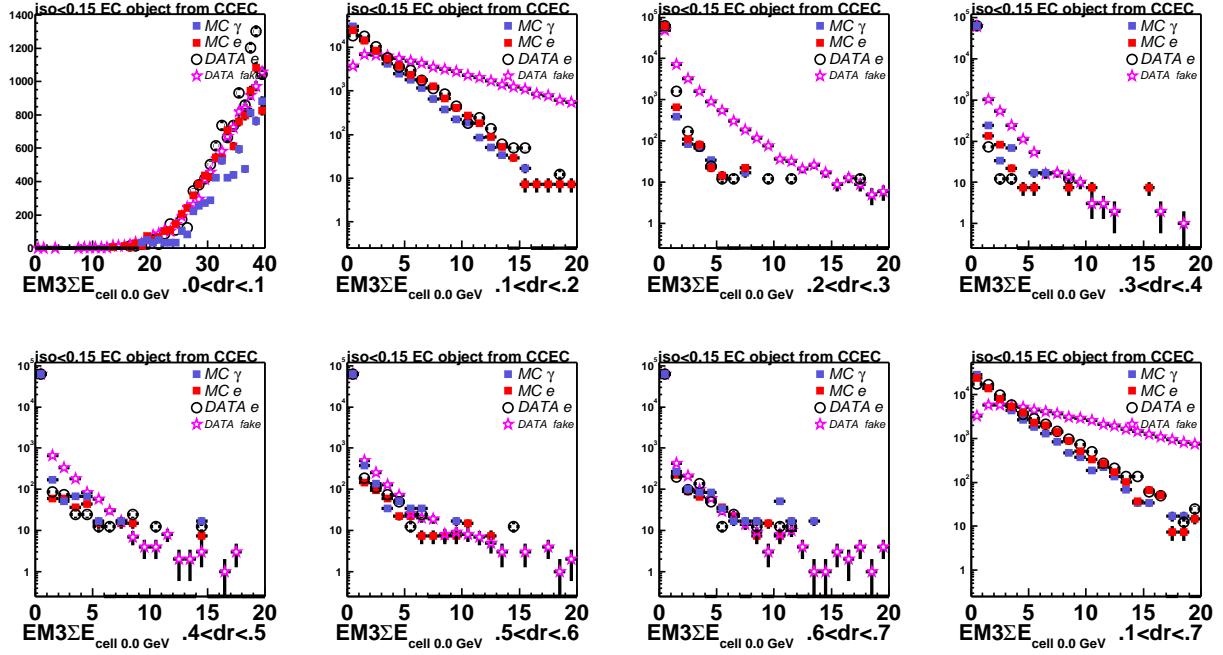


Figure 36: EC EM3: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

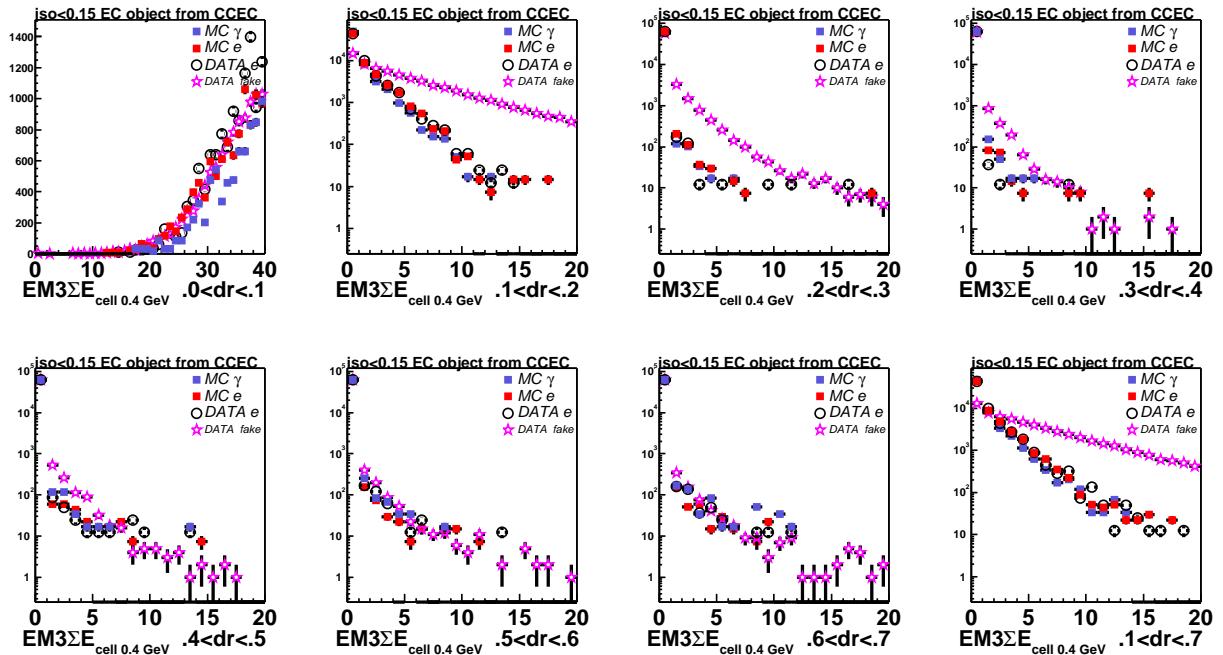


Figure 37: EC EM3: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

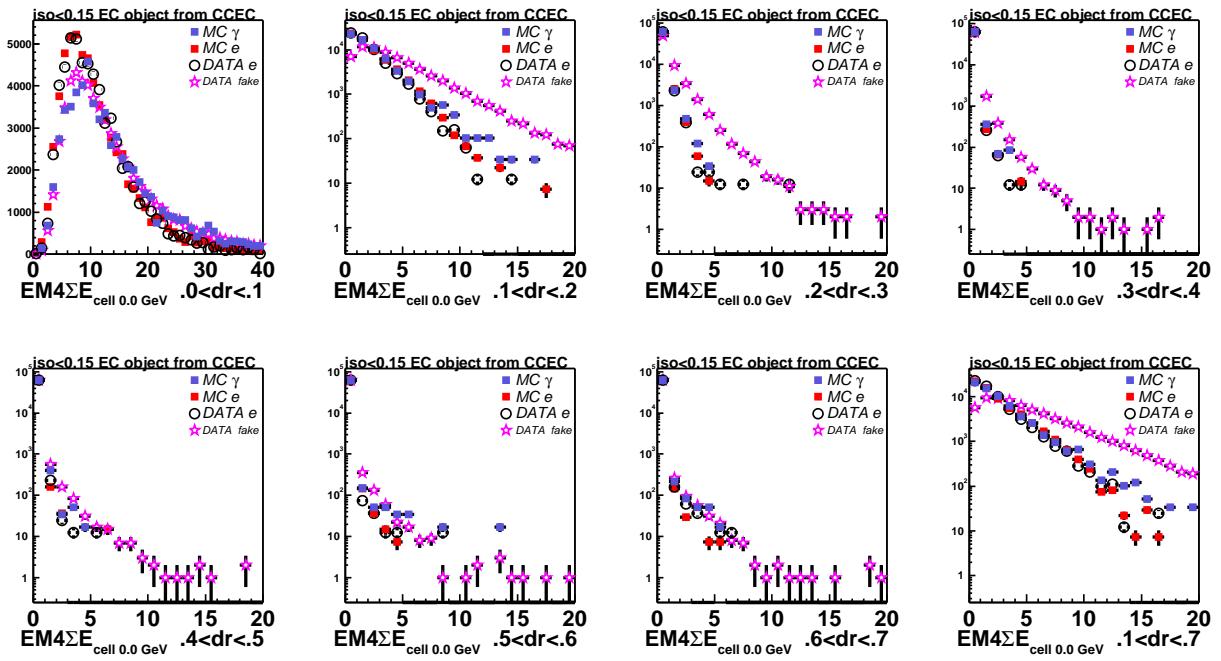


Figure 38: EC EM4: sum of cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

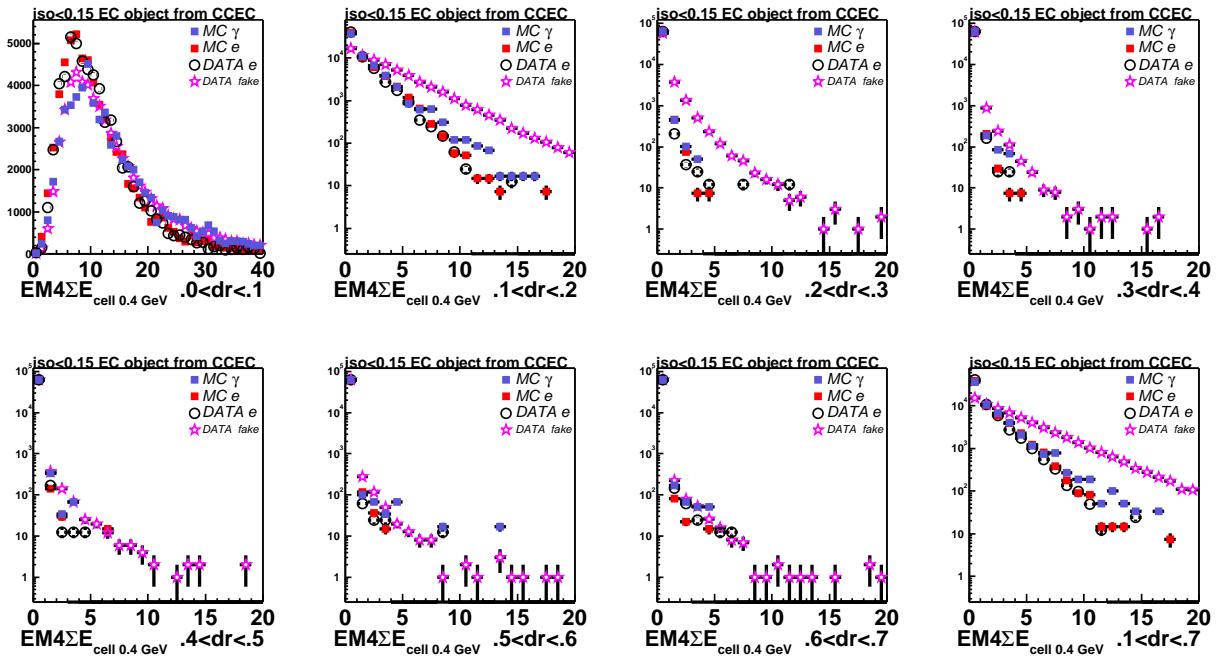


Figure 39: EC EM4: sum of cell energies in dR rings, (cell energy threshold is 400 MeV).

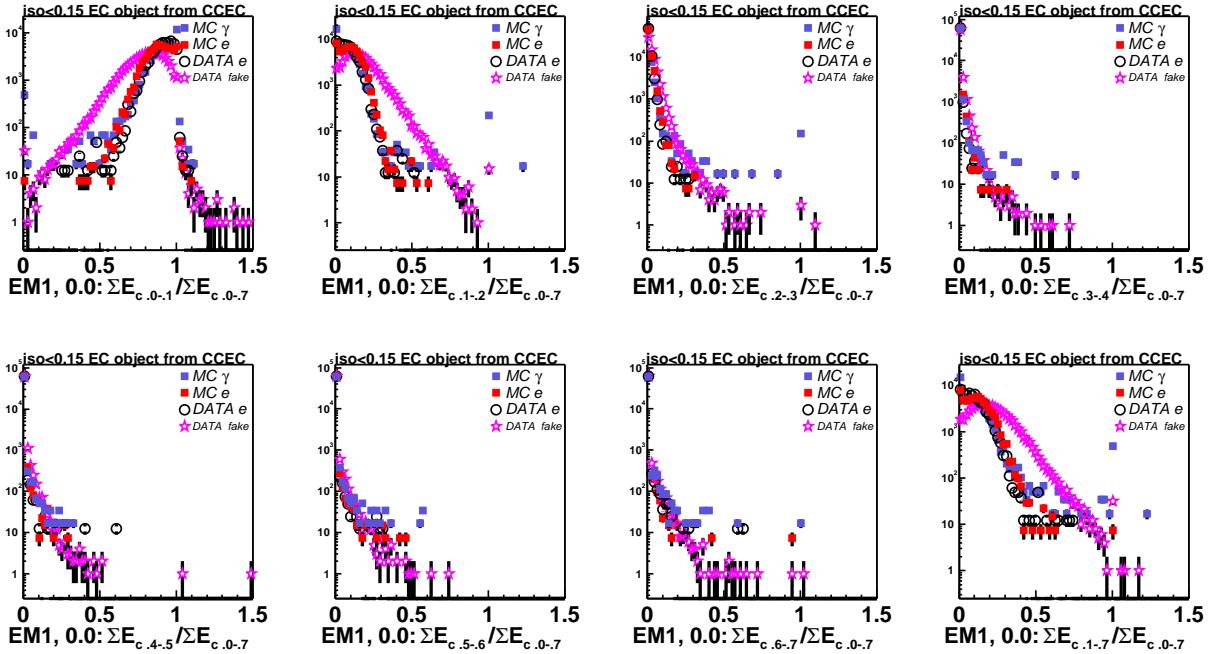


Figure 40: EC EM1: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

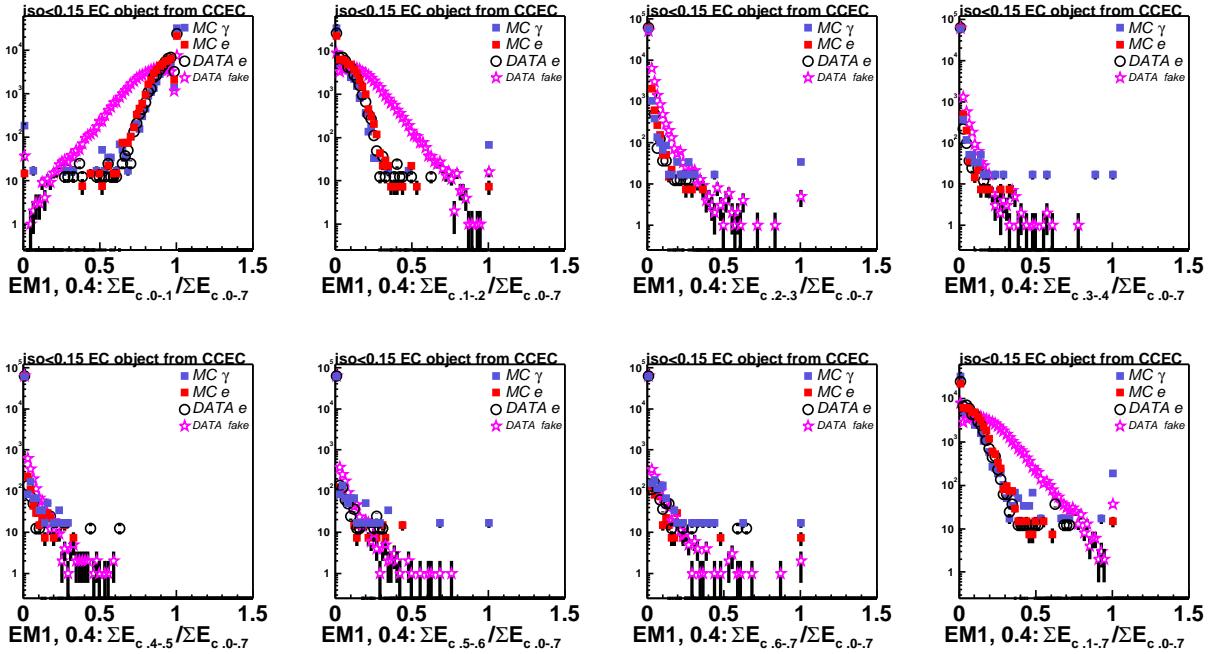


Figure 41: EC EM1: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

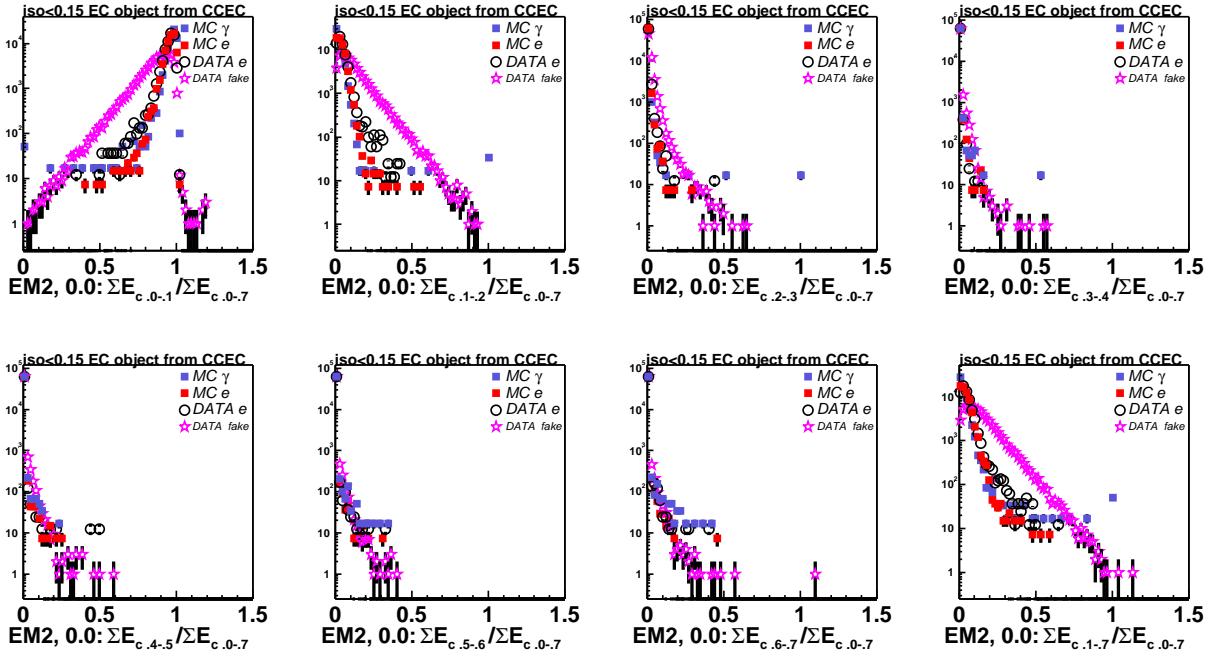


Figure 42: EC EM2: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

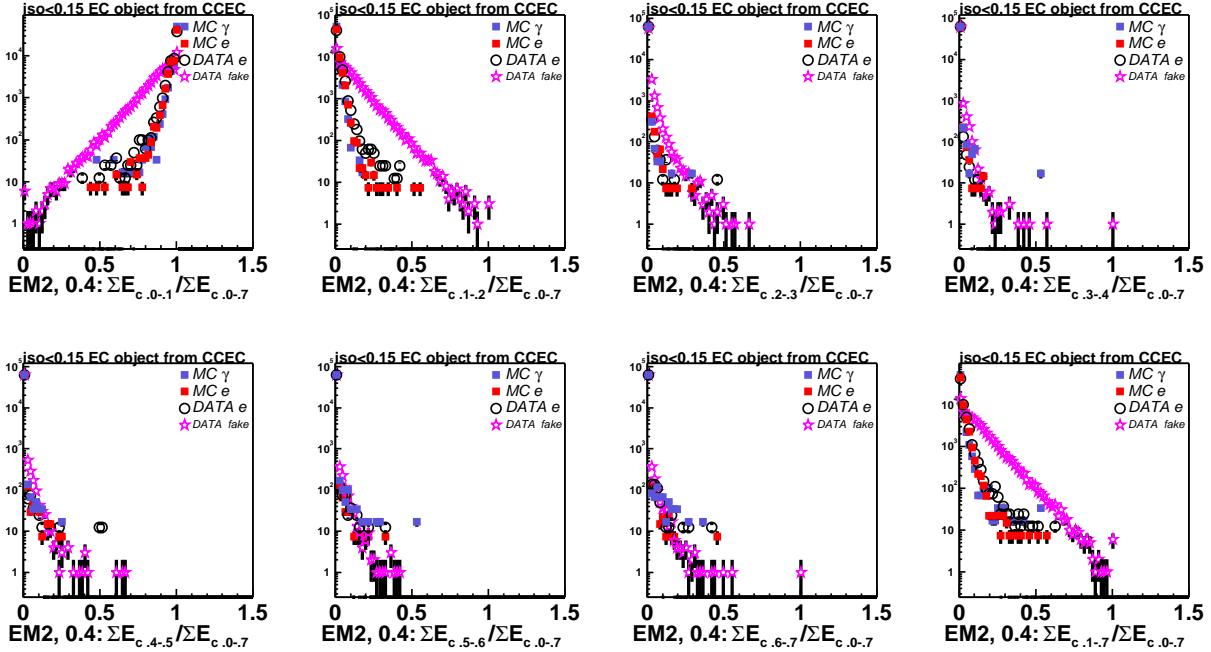


Figure 43: EC EM2: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

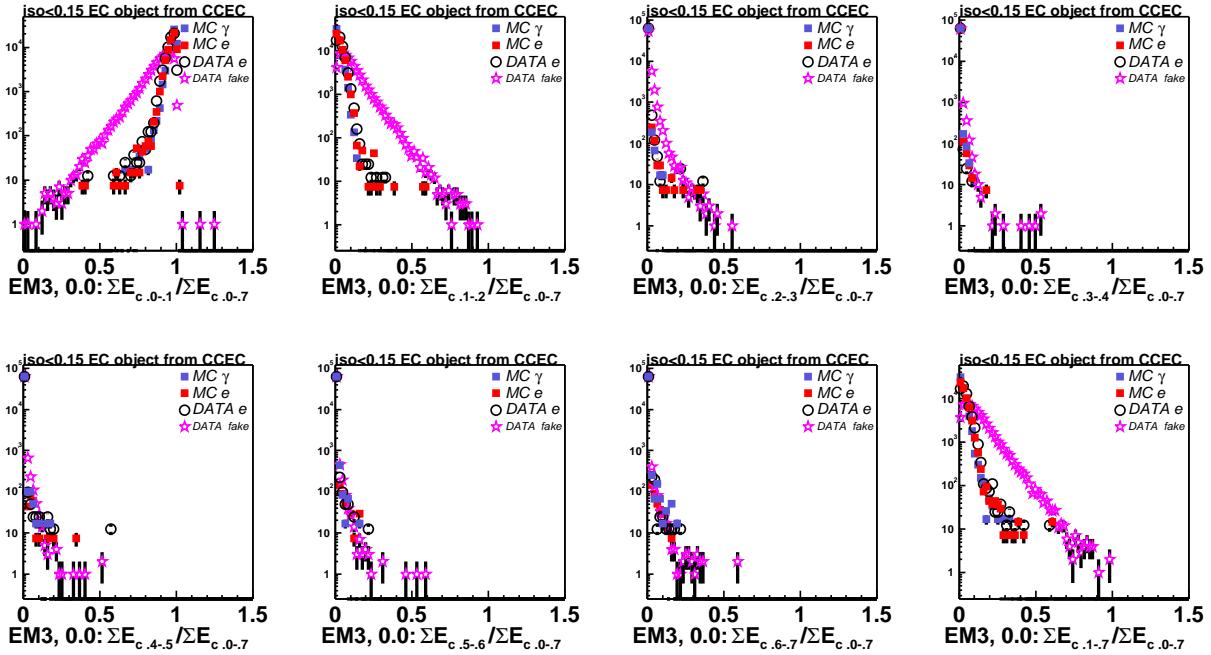


Figure 44: EC EM3: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

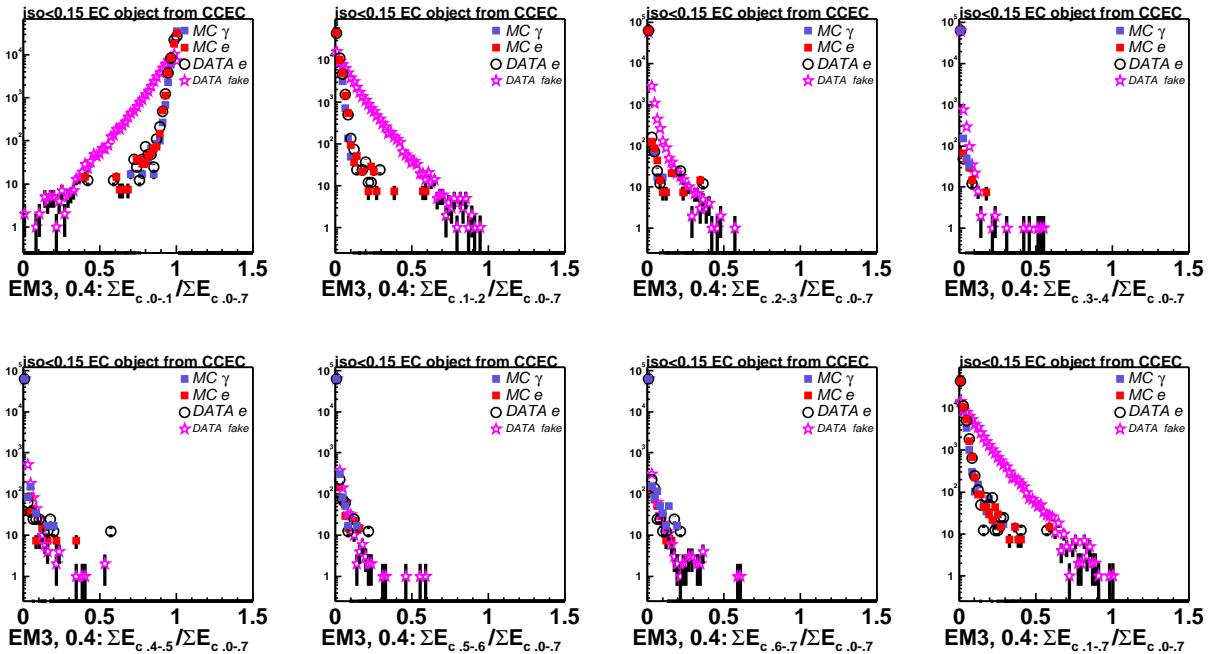


Figure 45: EC EM3: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

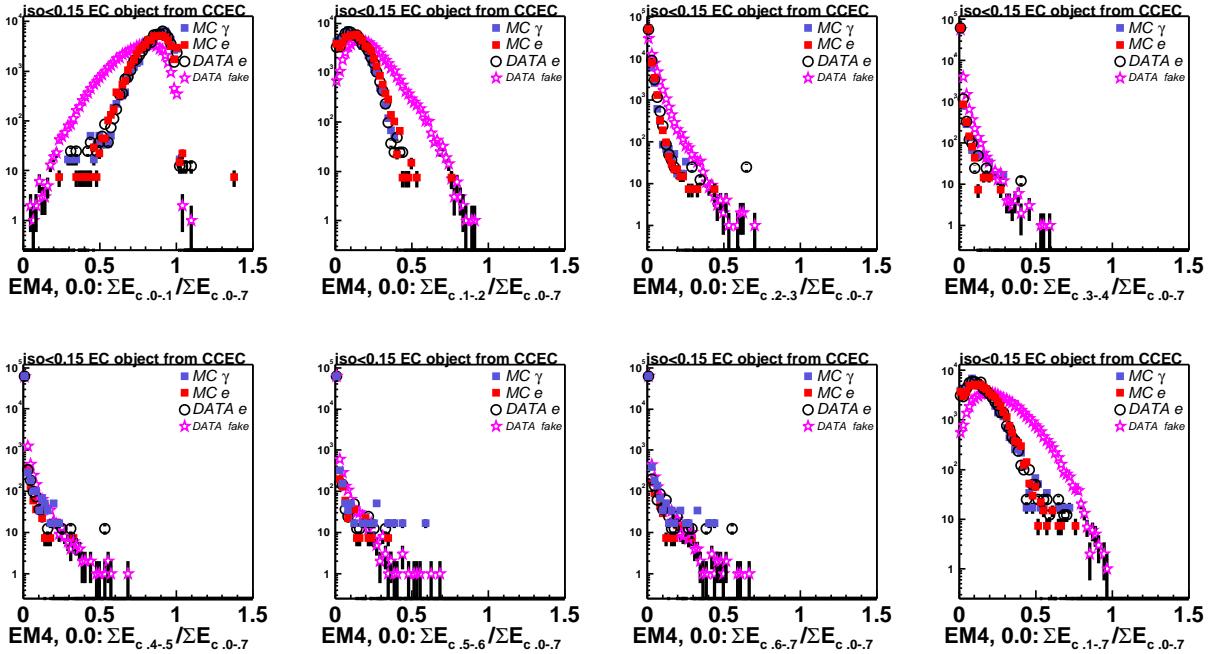


Figure 46: EC EM4: fractions of summed cell energies in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

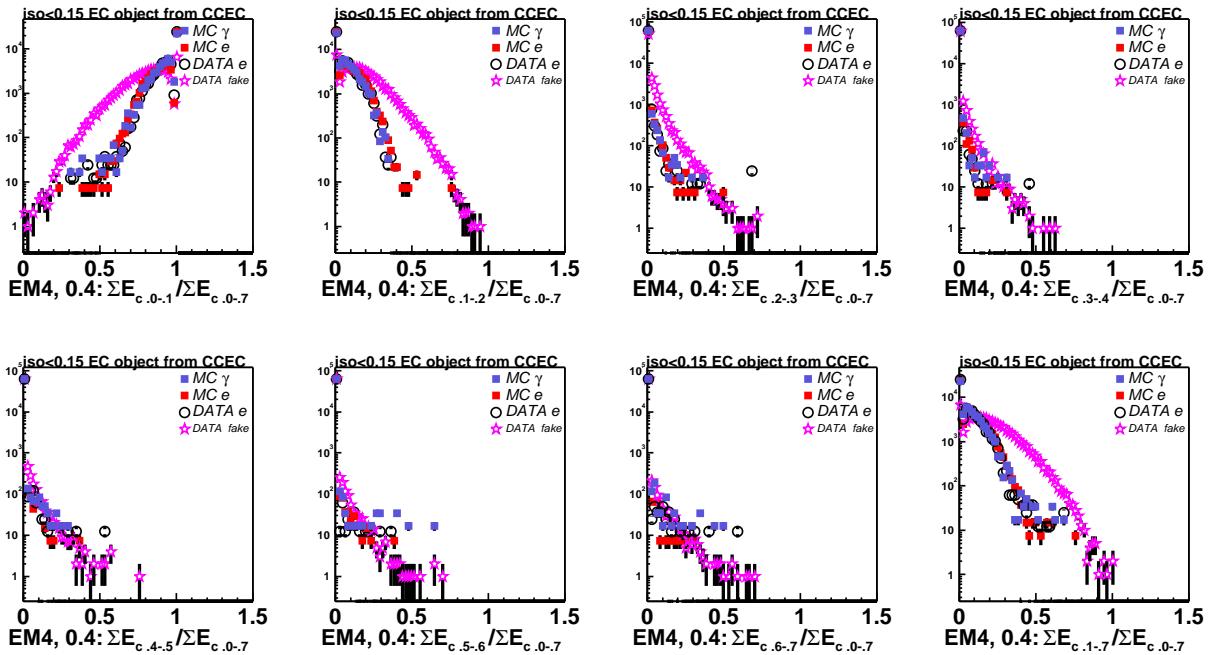


Figure 47: EC EM4: fractions of summed cell energies in dR rings, (cell energy threshold is 400 MeV).

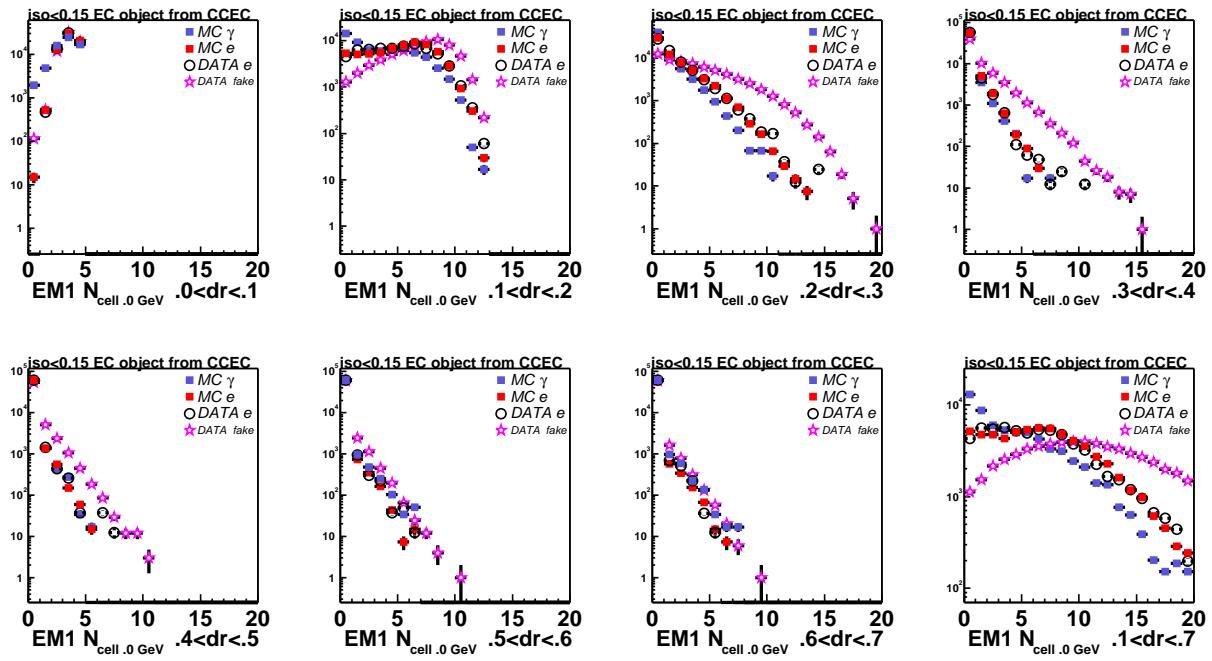


Figure 48: EC EM1: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

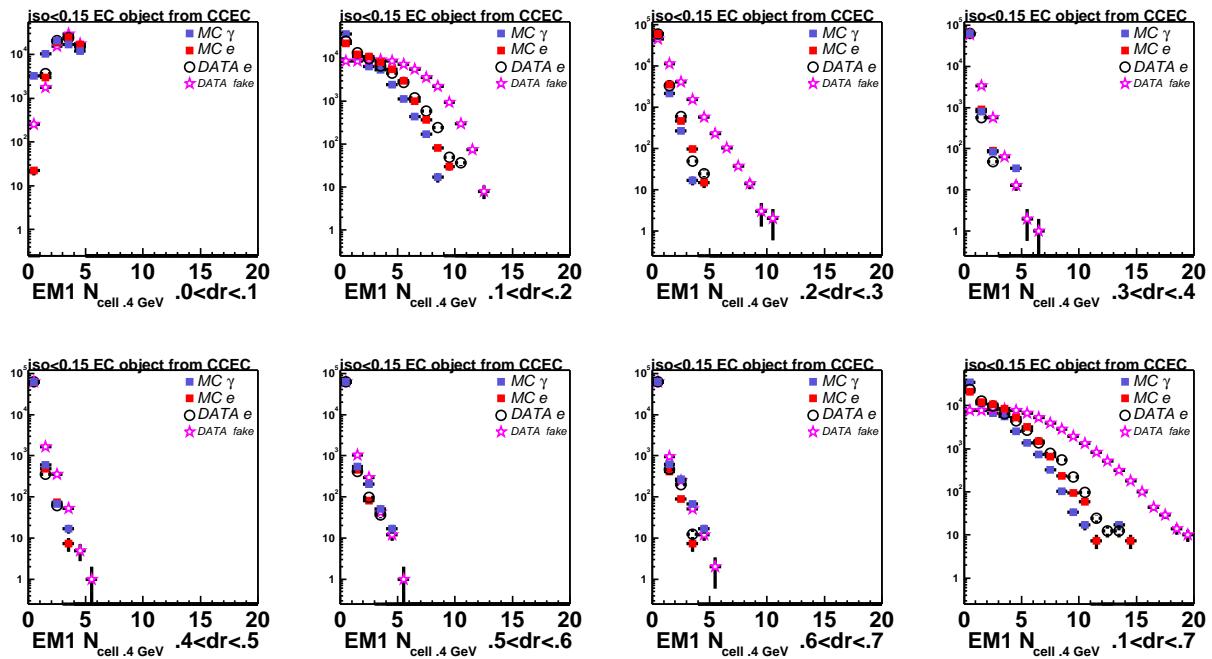


Figure 49: EC EM1: number of cells in dR rings, (cell energy threshold is 400 MeV).

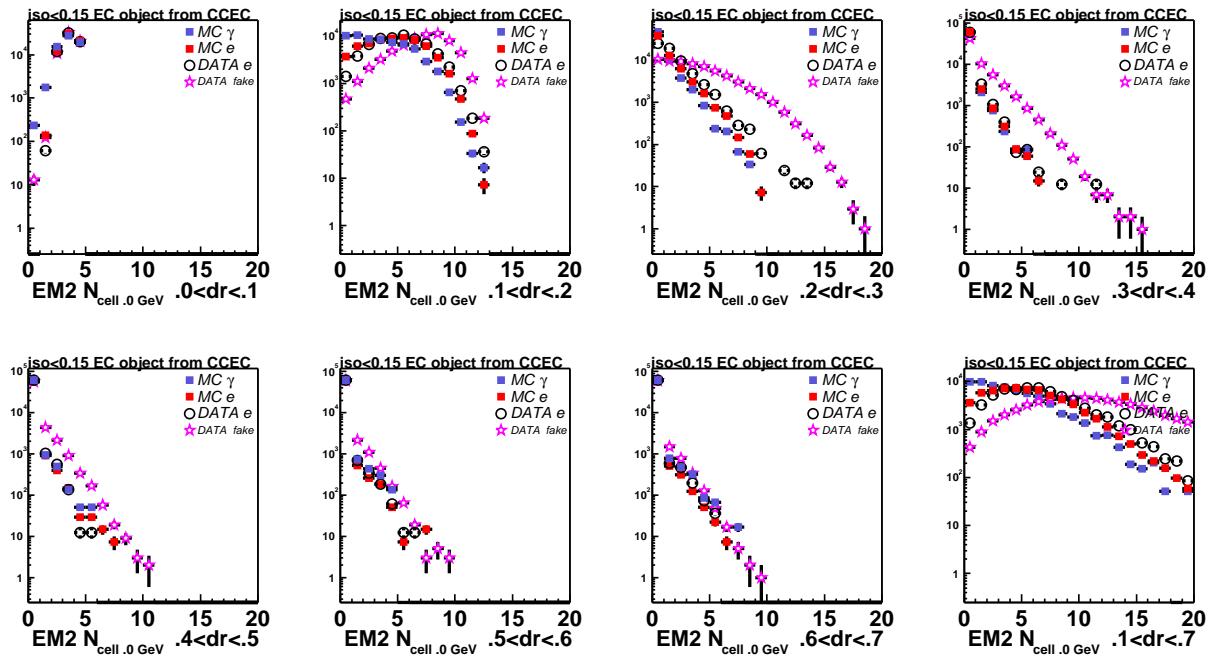


Figure 50: EC EM2: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

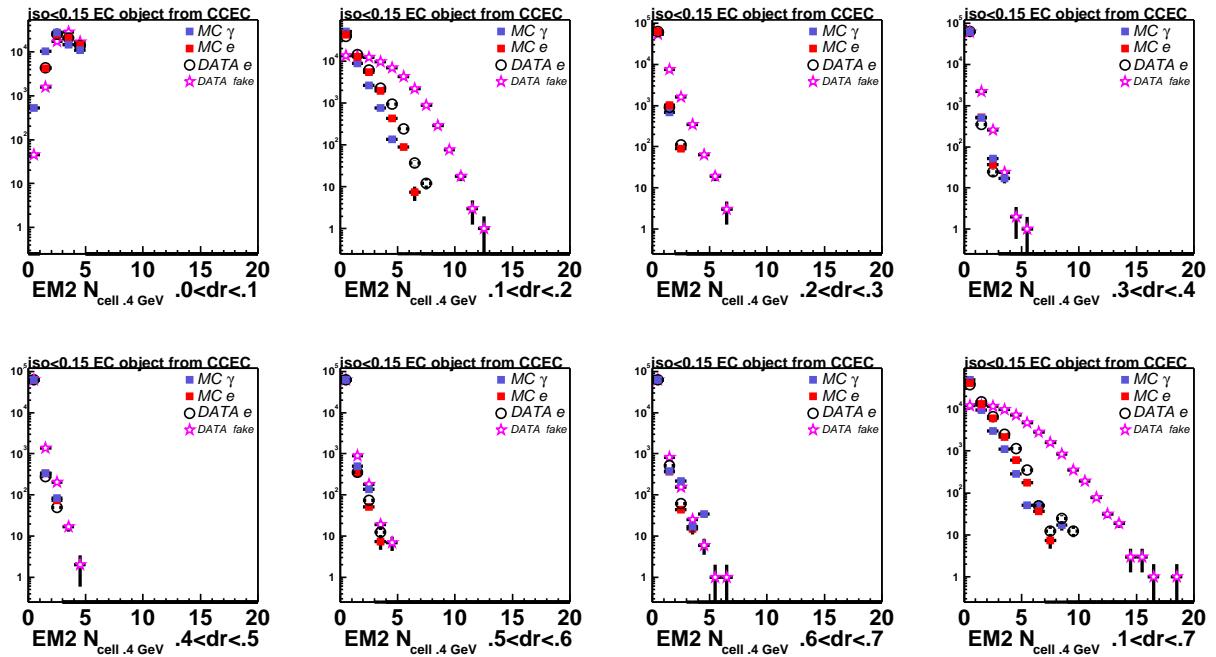


Figure 51: EC EM2: number of cells in dR rings, (cell energy threshold is 400 MeV).

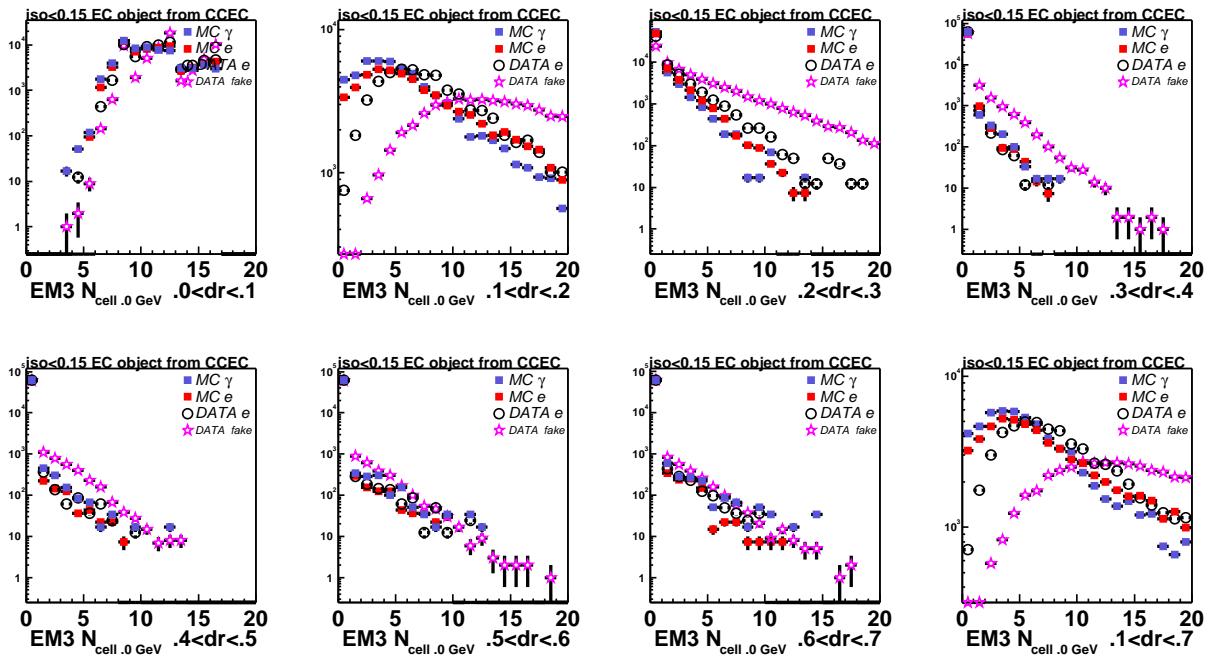


Figure 52: EC EM3: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

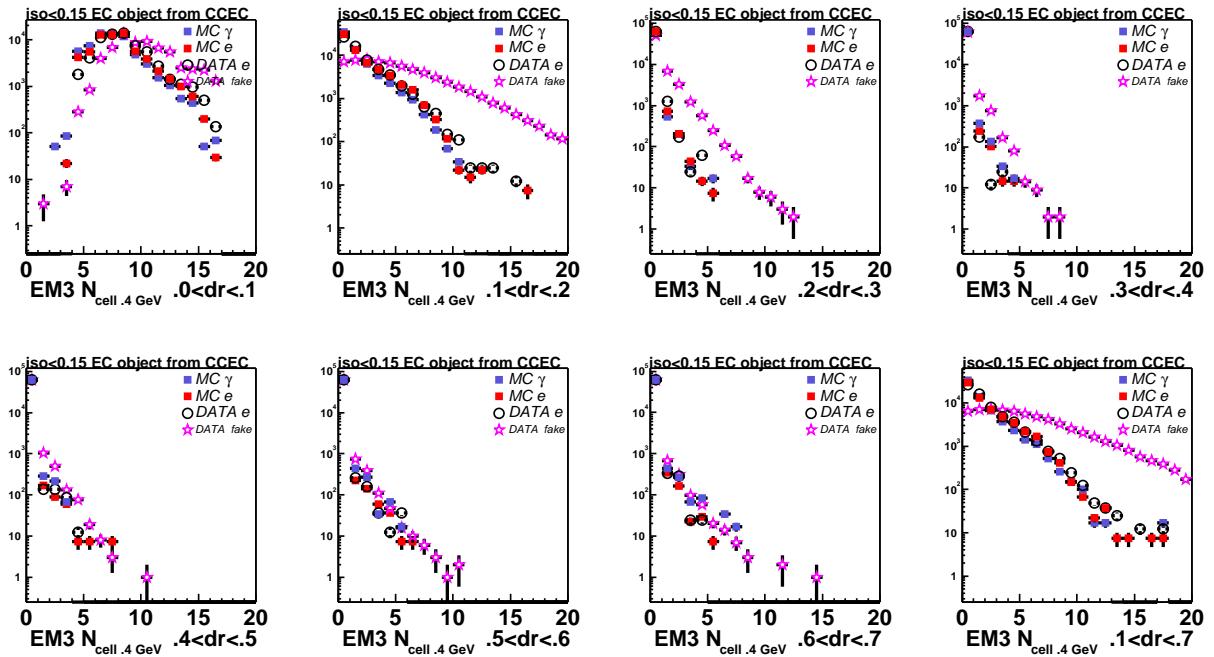


Figure 53: EC EM3: number of cells in dR rings, (cell energy threshold is 400 MeV).

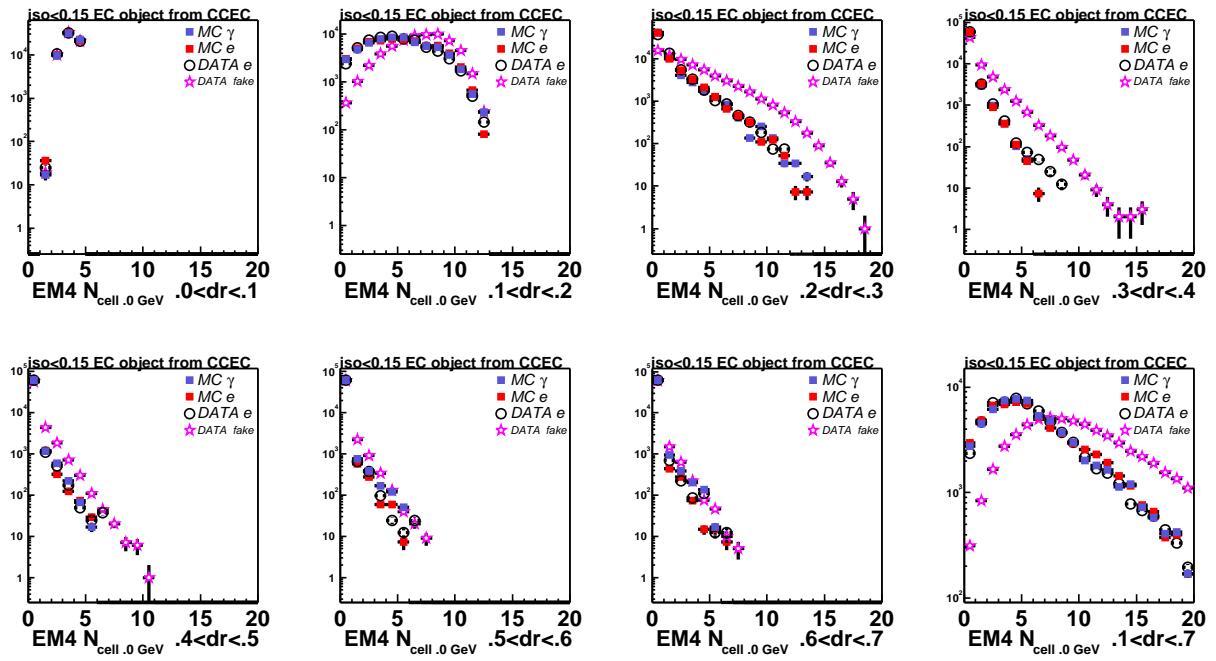


Figure 54: EC EM4: number of cells in dR rings, (no cell energy threshold at the analysis stage, 100(?) MeV in emreco).

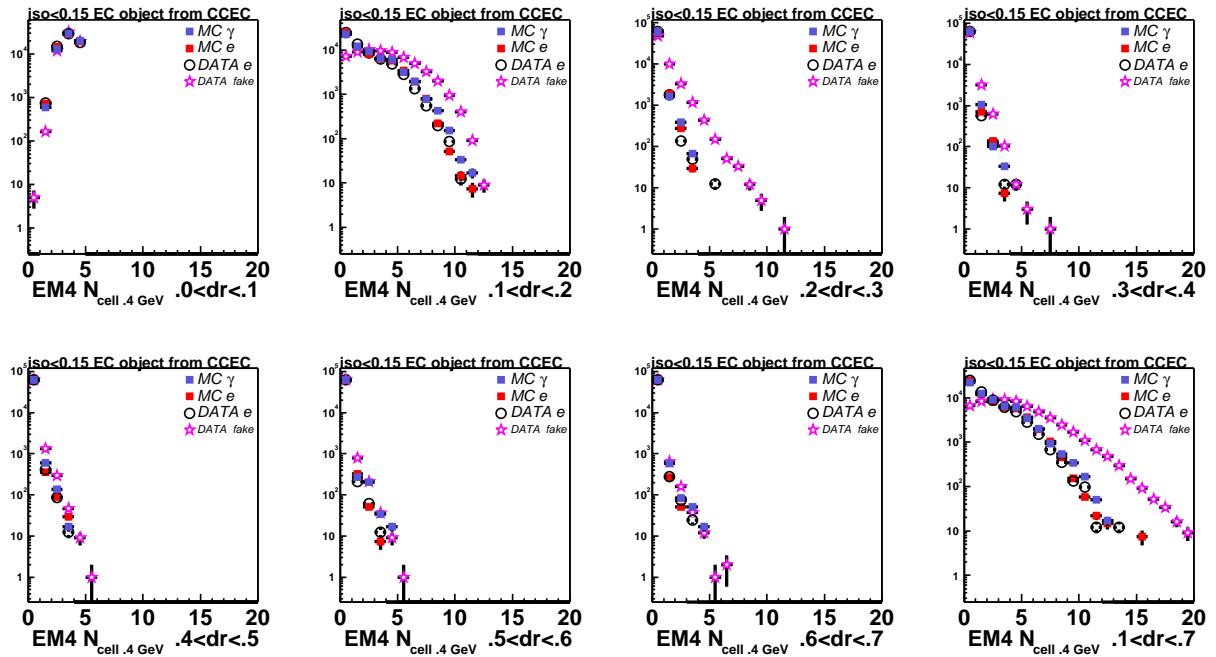


Figure 55: EC EM4: number of cells in dR rings, (cell energy threshold is 400 MeV).

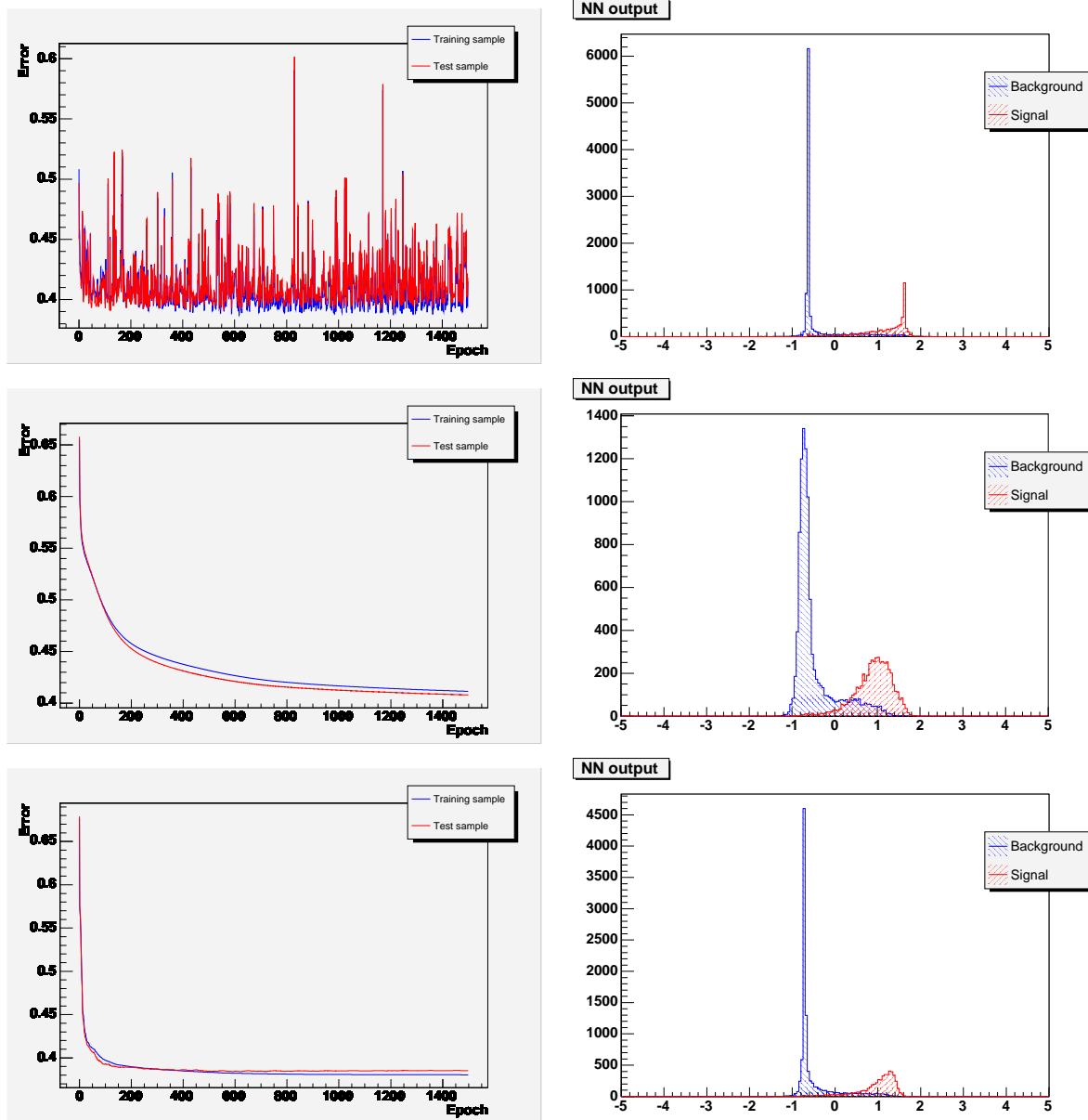


Figure 56: NN training curves and NN output in CC with three learning methods. Left column: error for training and test sample as a function of the number of epochs (total number of training epochs for each method is 1500). Right column: NN output. 7 NN variables are used: those in HMx7 minus $Z(\text{primary vertex})$ (EM3 $r\phi$ -width, four EM floor energy fractions, $\log(E)$) plus track isolation. One hidden layer of 14 hidden nodes is used. Top row: stochastic, middle row: steepest descent, bottom row: BFGS.

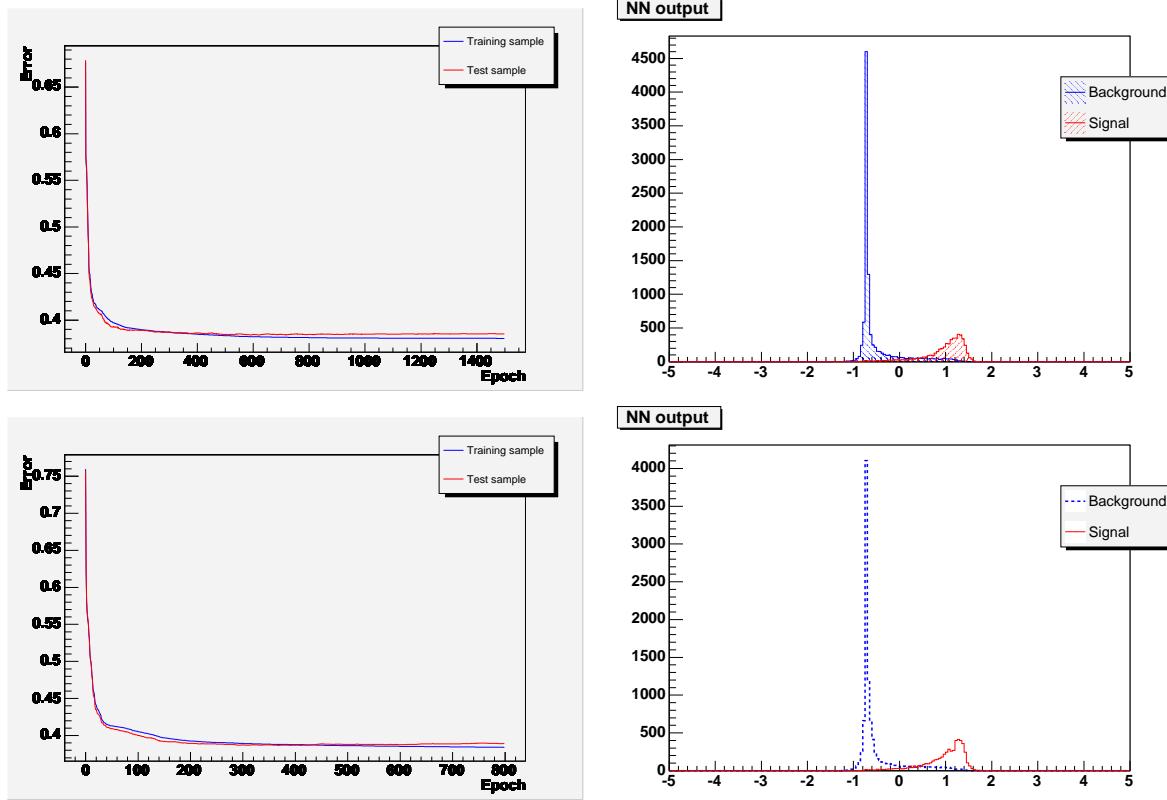


Figure 57: NN training curves and NN output in CC. Left column: error for training and test sample as a function of the number of epochs. Right column: NN output. 7 NN variables are used: those in HMX7 minus Z(primary vertex) (EM3 $r\phi$ -width, four EM floor energy fractions, $\log(E)$) plus track isolation. One hidden layer of 14 hidden nodes is used. Top row: 1500 epochs, bottom row: 800 epochs.

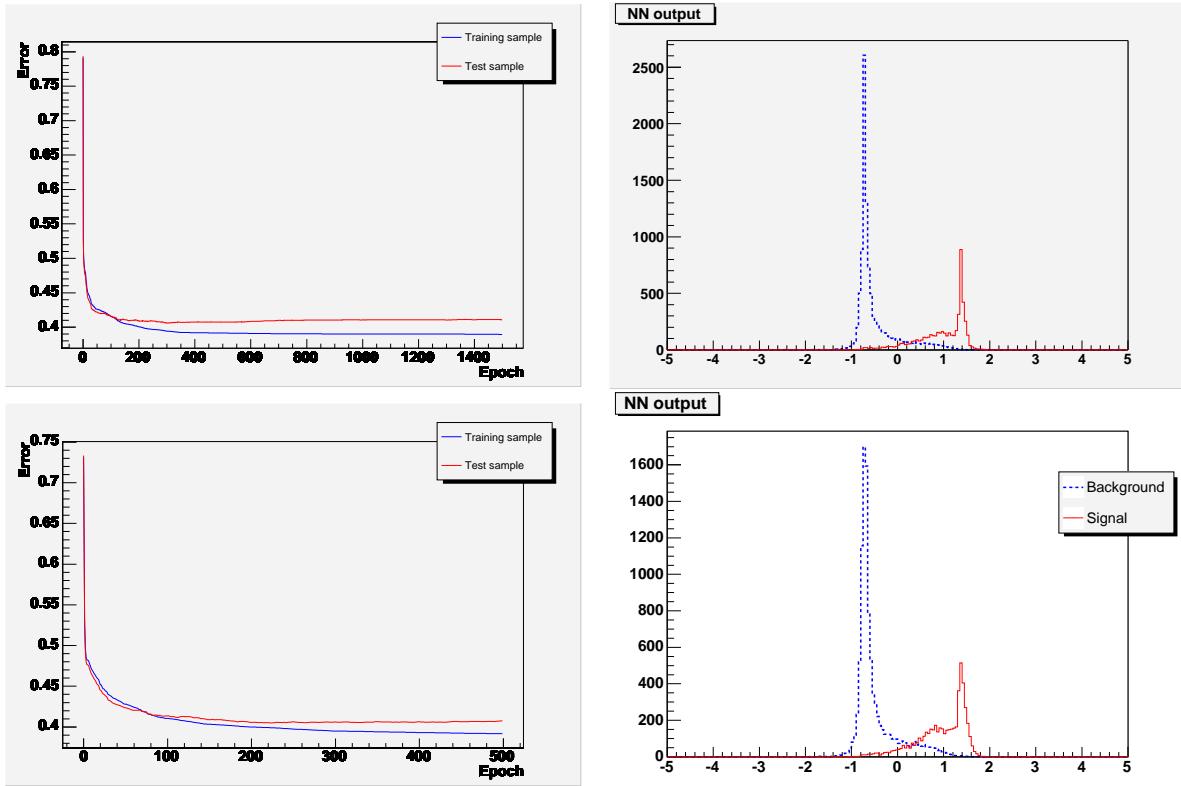


Figure 58: NN training curves and NN output in EC. Left column: error for training and test sample as a function of the number of epochs. Right column: NN output. 8 NN variables are used: those in HMx8 minus Z(primary vertex) (EM3 $r\phi$ -width, EM3 r-width, four EM floor energy fractions, $\log(E)$) plus track isolation. One hidden layer of 15 hidden nodes is used. Top row: 1500 epochs, bottom row: 500 epochs.

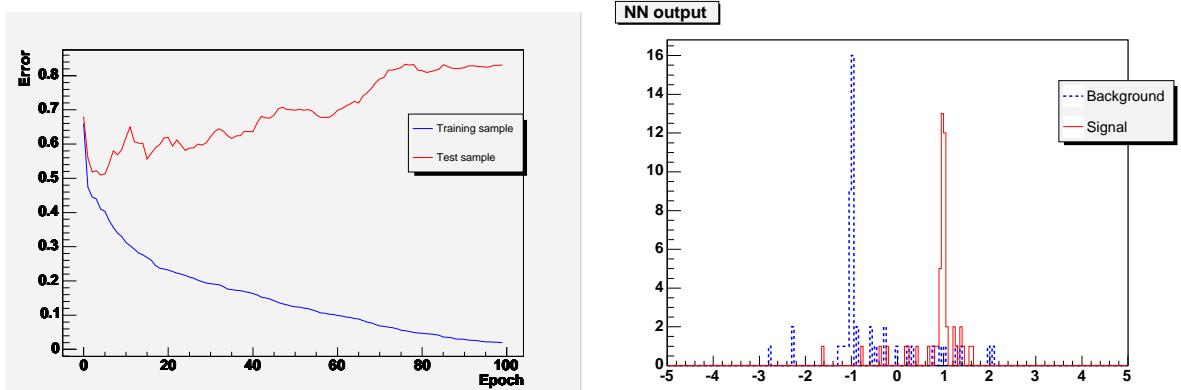


Figure 59: An example of overtrained NN for the input sample of 50 events. Left: error for training and test sample as a function of the number of epochs. Right: NN output. 8 NN variables are used: those in HMx8 minus Z(primary vertex) (EM3 $r\phi$ -width, EM3 r-width, four EM floor energy fractions, $\log(E)$) plus track isolation. One hidden layer of 15 hidden nodes is used.

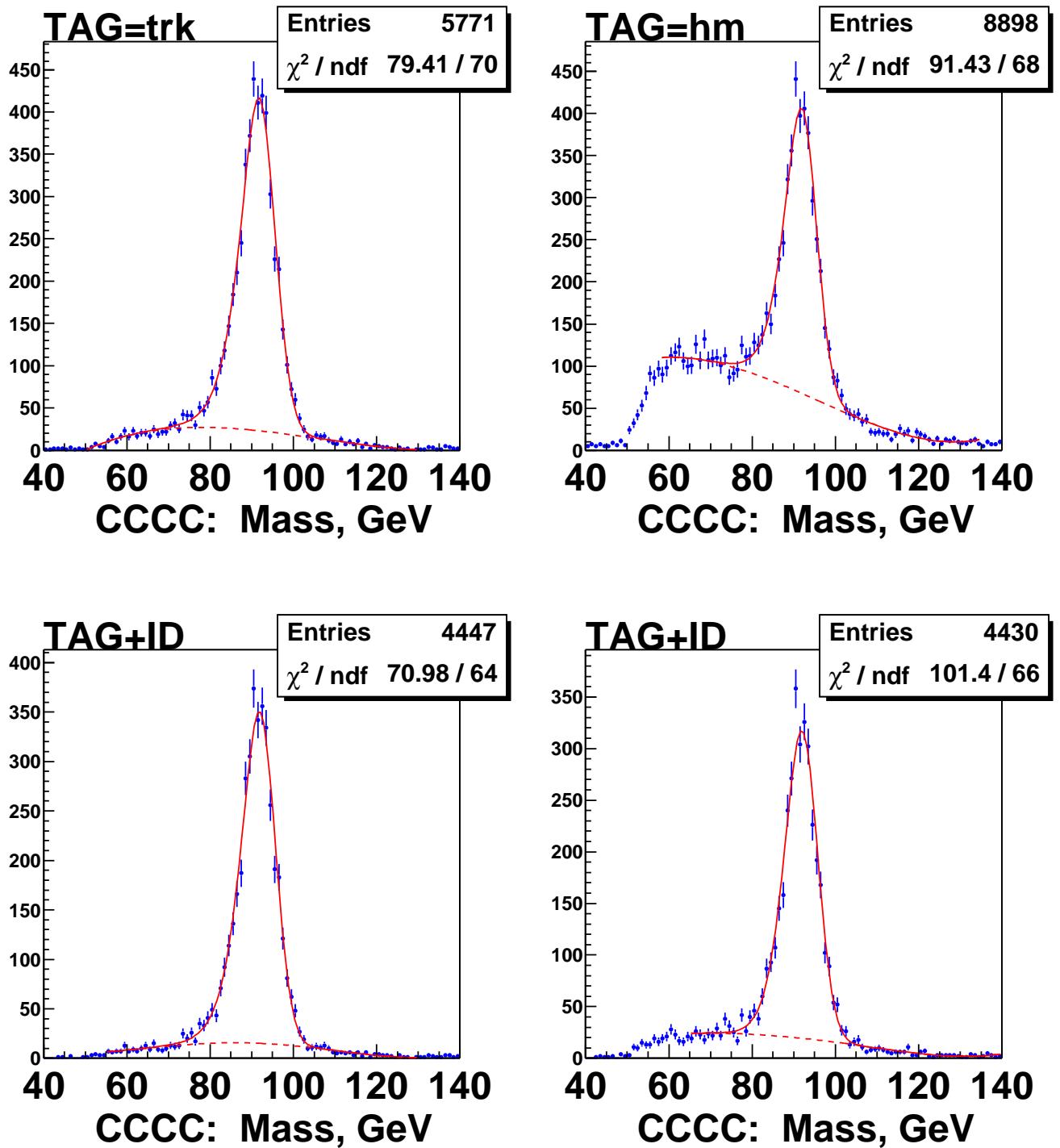


Figure 60: EMID in CC = HMx7<10 plus Track Isolation <2 GeV. CCCC diEM mass distributions for the EMID efficiency measurement in CC are shown. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($\text{HM8} : \chi^2 < 5$) is applied to the tag object. Top plots shows the denominator mass distributions i.e. when a cut is applied to the tag object only. Bottom plots show numerator mass distributions i.e. when EMID cut is applied to the probe object.

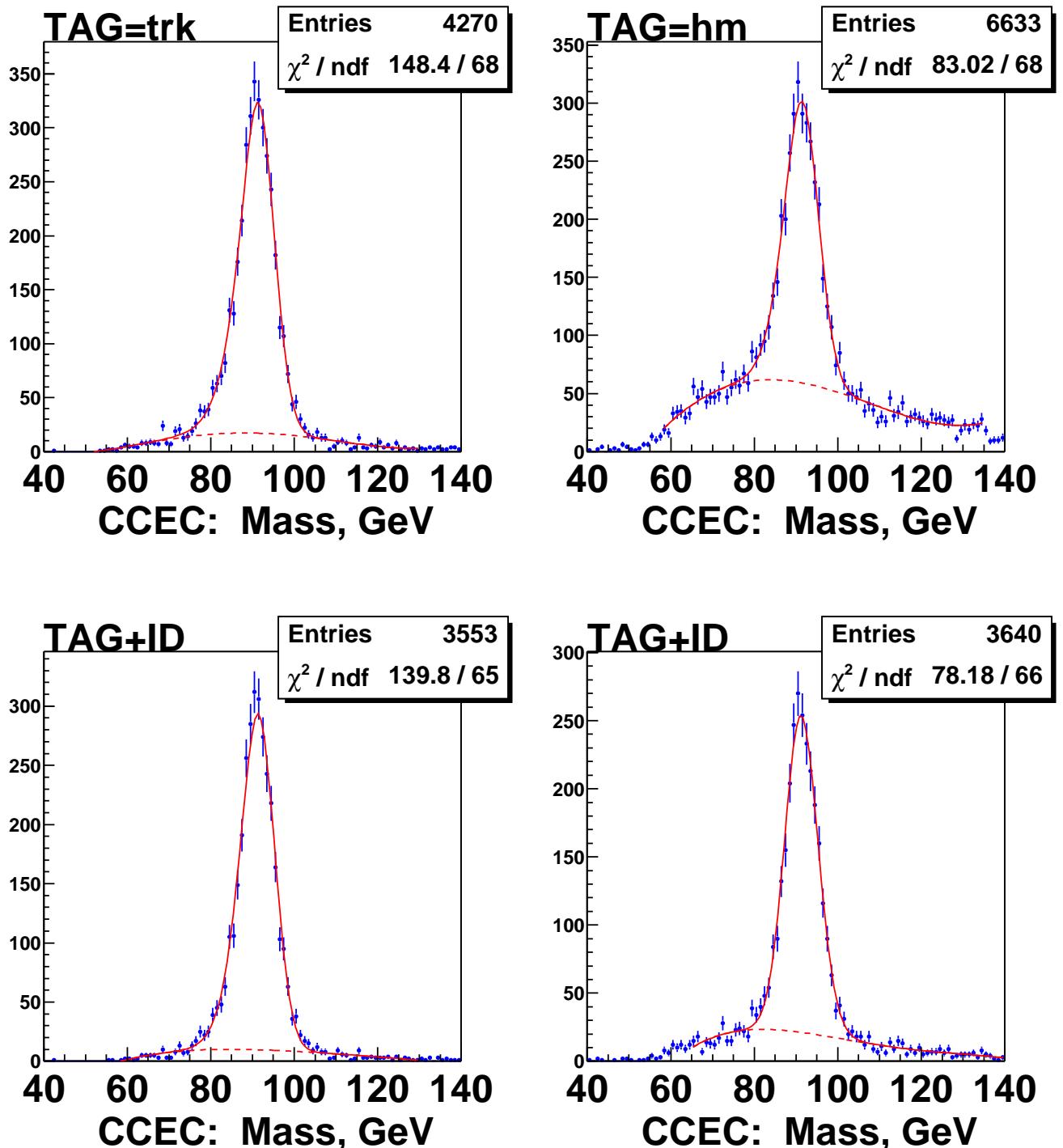


Figure 61: EMID in EC = HM_x8<15 plus Track Isolation <2 GeV. CCEC diEM mass distributions for the EMID efficiency measurement in EC are shown. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut (HM8 : $\chi^2 < 5$) is applied to the tag object. Top plots shows the denominator mass distributions i.e. when a cut is applied to the tag object only. Bottom plots show numerator mass distributions i.e. when EMID cut is applied to the probe object.

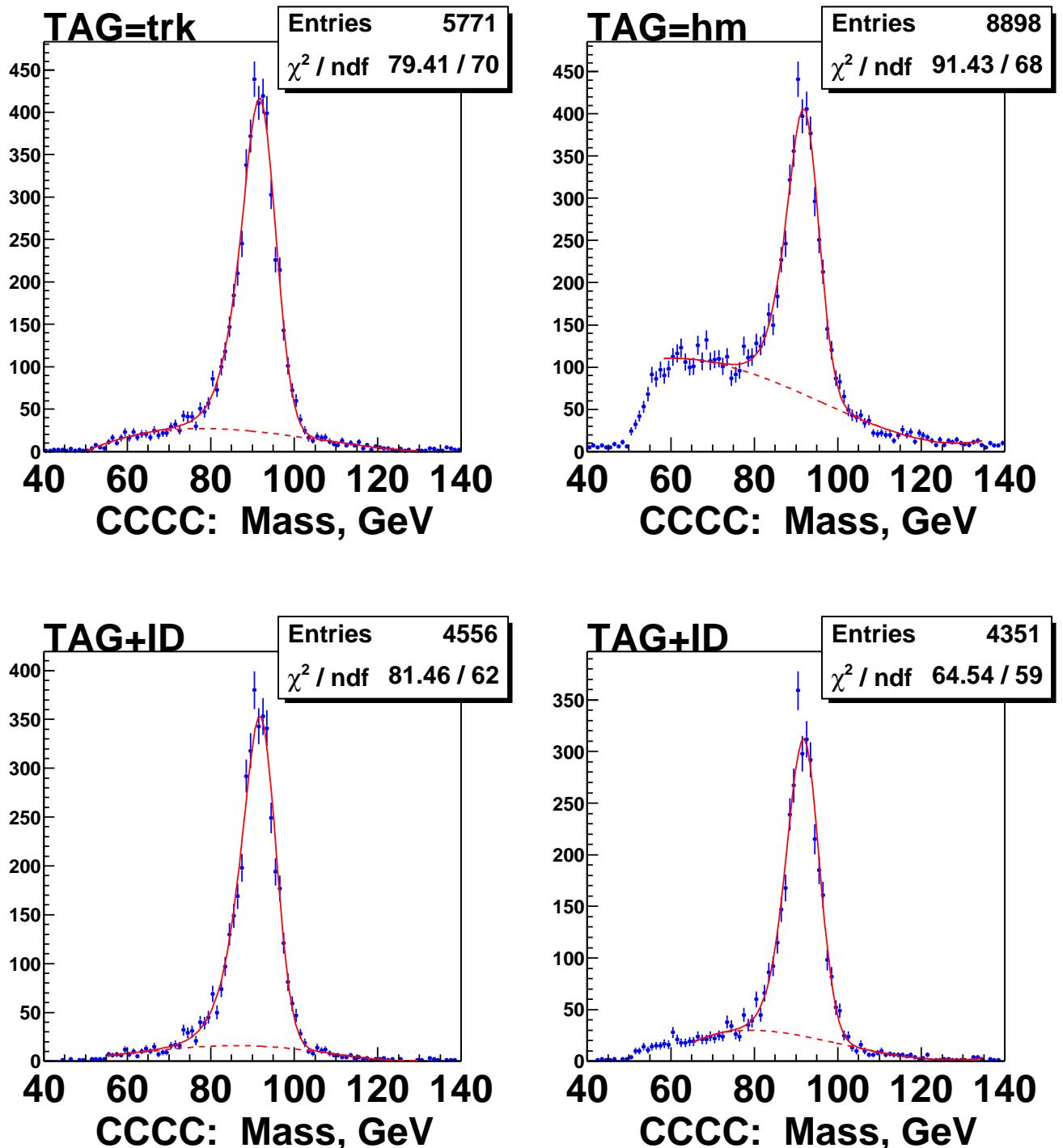


Figure 62: CCCC diEM mass distributions for the CC EMID efficiency measurement. EMID in CC = Neural Net output > 0.1 . Training was done with 1500 epochs. 7 NN variables are used: those in HMx7 minus Z(primary vertex) (EM3 rphi-width, four EM floor energy fractions, $\log(E)$) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : \chi^2 < 5$) is applied to the tag object. Top plots show the denominator mass distributions i.e. when a cut is applied to the tag object only. Bottom plots show numerator mass distributions i.e. when EMID cut is applied to the probe object.

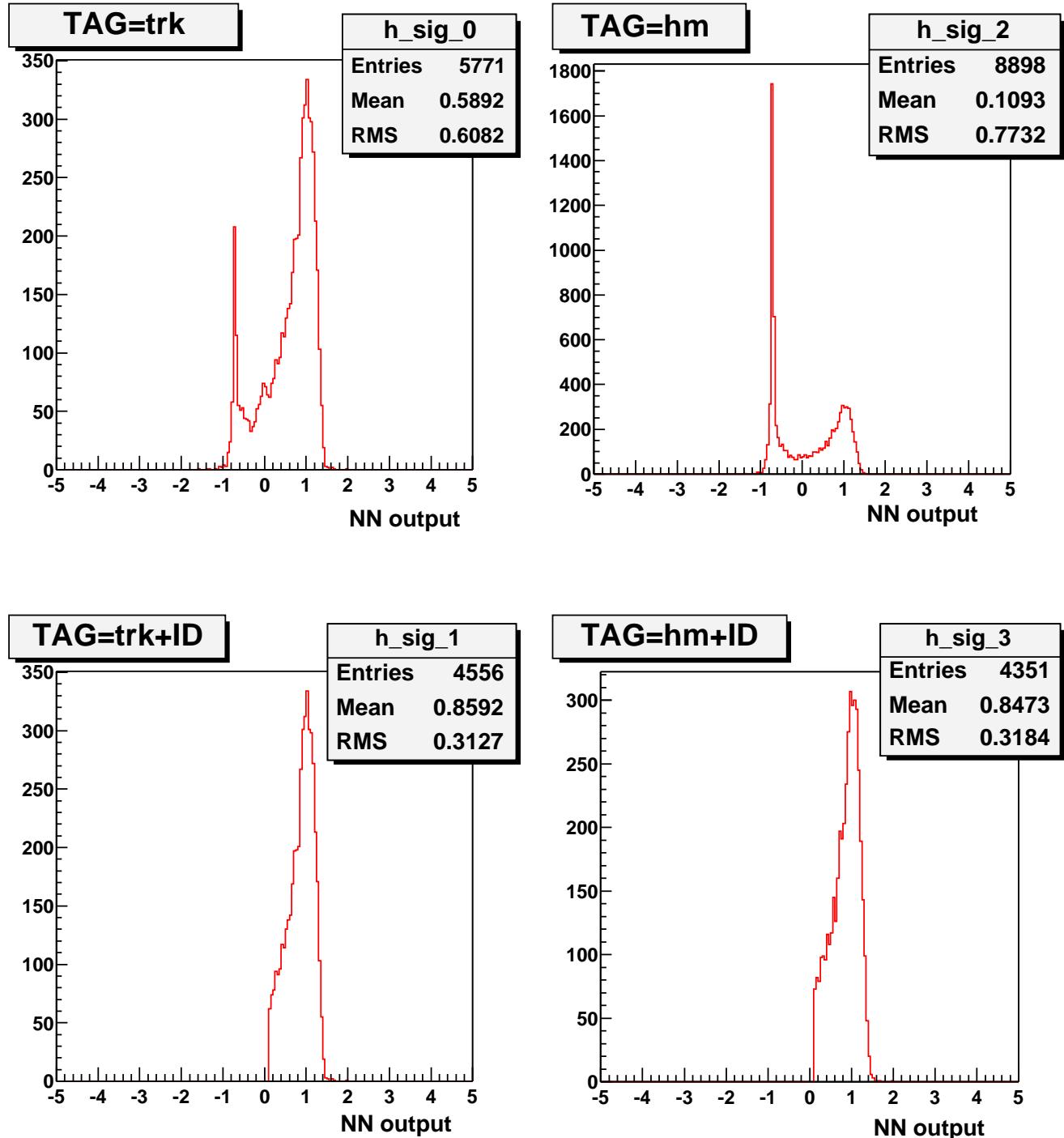


Figure 63: NN output for the probe object corresponding to the CC EMID efficiency measurement. EMID in CC = Neural Net output > 0.1 . Training was done with 1500 epochs. 7 NN variables are used: those in HMx7 minus Z(primary vertex) (EM3 rphi-width, four EM floor energy fractions, $\log(E)$) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : chi^2 < 5$) is applied to the tag object. Top plots shows the denominator NN output i.e. when no cut is applied to the probe object. Bottom plots show numerator NN output when EMID cut is applied to the probe object.

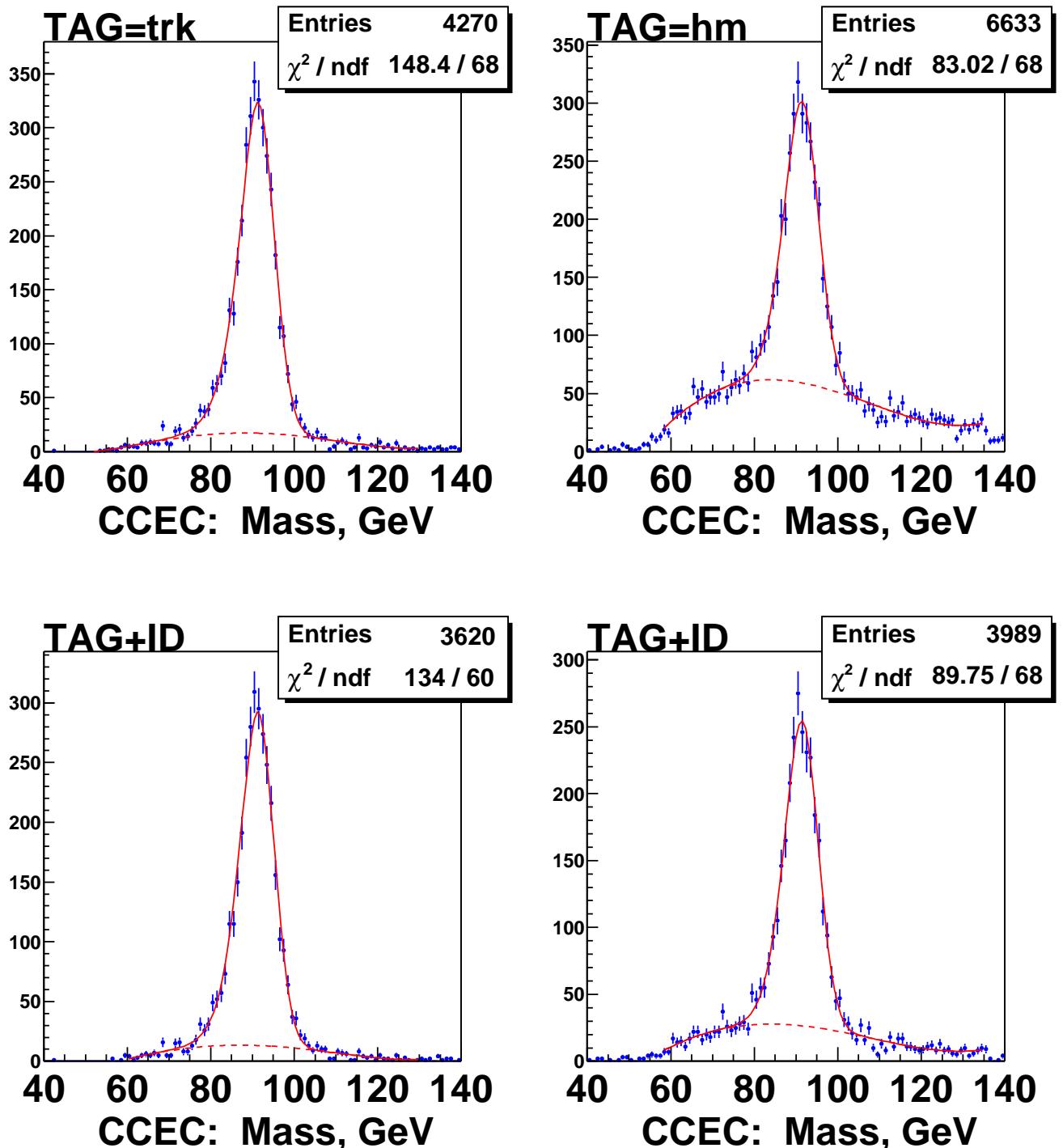


Figure 64: CCEC diEM mass distributions for the EC EMID efficiency measurement. EMID in EC = Neural Net output > -0.1 . Training was done with 1500 epochs. 8 NN variables are used: those in HMx8 minus Z(primary vertex) (EM3 rphi-width, four EM floor energy fractions, $\log(E)$) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : \chi^2 < 5$) is applied to the tag object. Top plots shows the denominator mass distributions i.e. when a cut is applied to the tag object only. Bottom plots show numerator mass distributions i.e. when EMID cut is applied to the probe object.

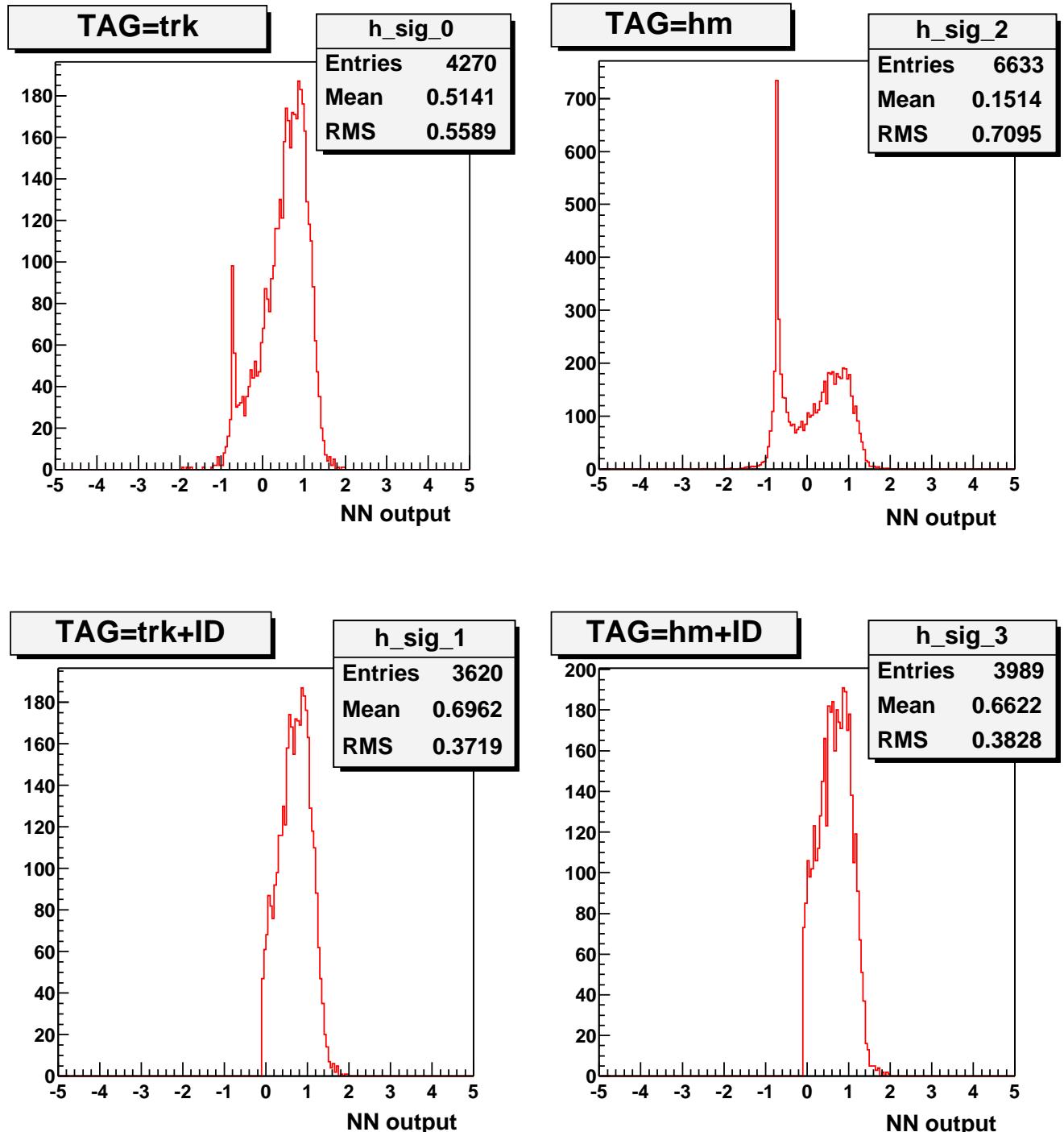


Figure 65: NN output for the probe object corresponding to the EC EMID efficiency measurement. EMID in EC = Neural Net output > -0.1 . Training was done with 1500 epochs. 8 NN variables are used: those in HMx8 minus Z(primary vertex) (EM3 rphi-width, four EM floor energy fractions, $\log(E)$) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : chi^2 < 5$) is applied to the tag object. Top plots shows the denominator NN output i.e. when no cut is applied to the probe object. Bottom plots show numerator NN output when EMID cut is applied to the probe object.

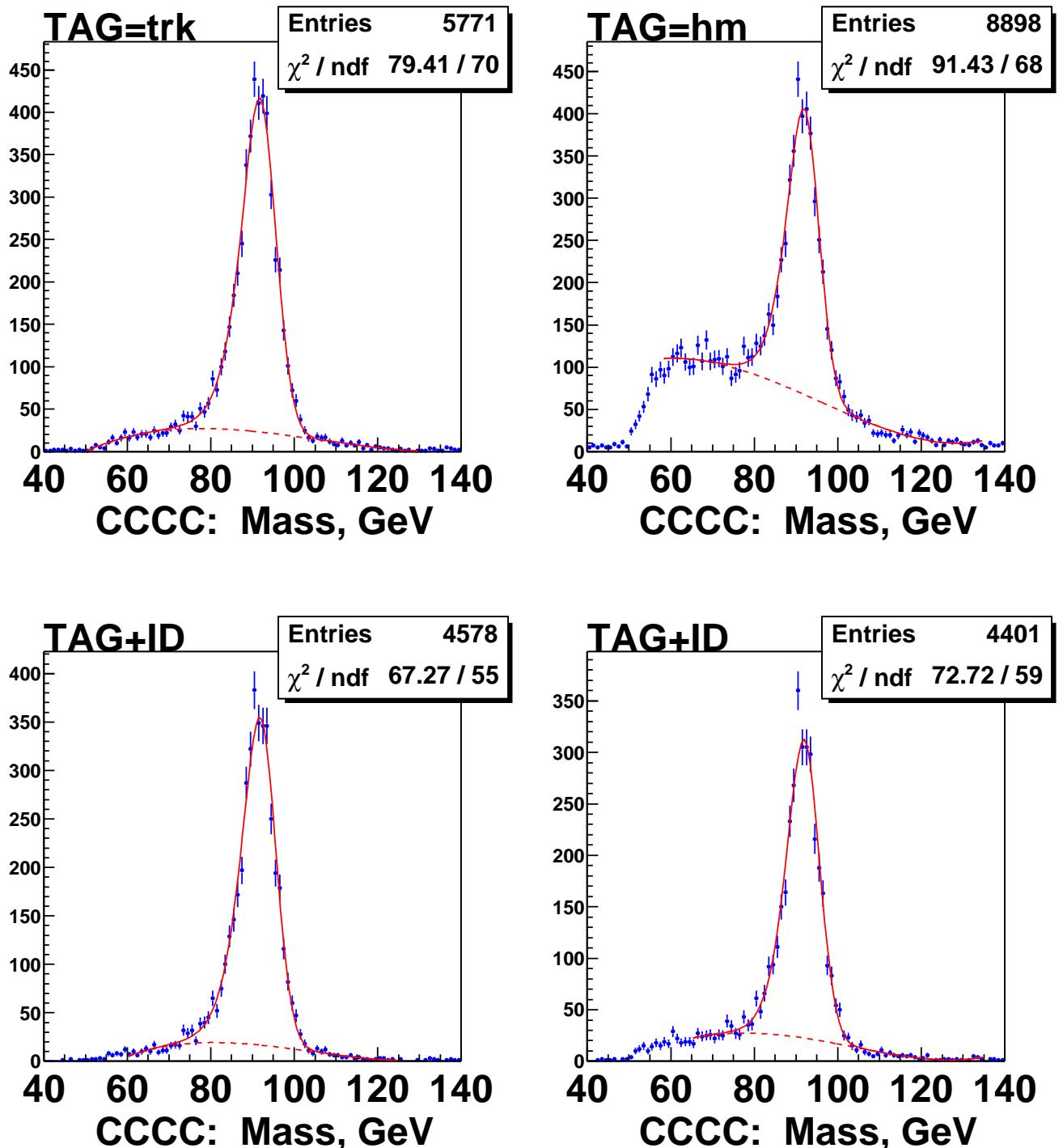


Figure 66: CCCC diEM mass distributions for the CC EMID efficiency measurement. EMID in CC = Neural Net output > 0.1 . Training was done with 800 epochs. 7 NN variables are used: those in HMx7 minus Z(primary vertex) (EM3 rphi-width, four EM floor energy fractions, $\log(E)$) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : chi^2 < 5$) is applied to the tag object. Top plots show the denominator mass distributions i.e. when a cut is applied to the tag object only. Bottom plots show numerator mass distributions i.e. when EMID cut is applied to the probe object.

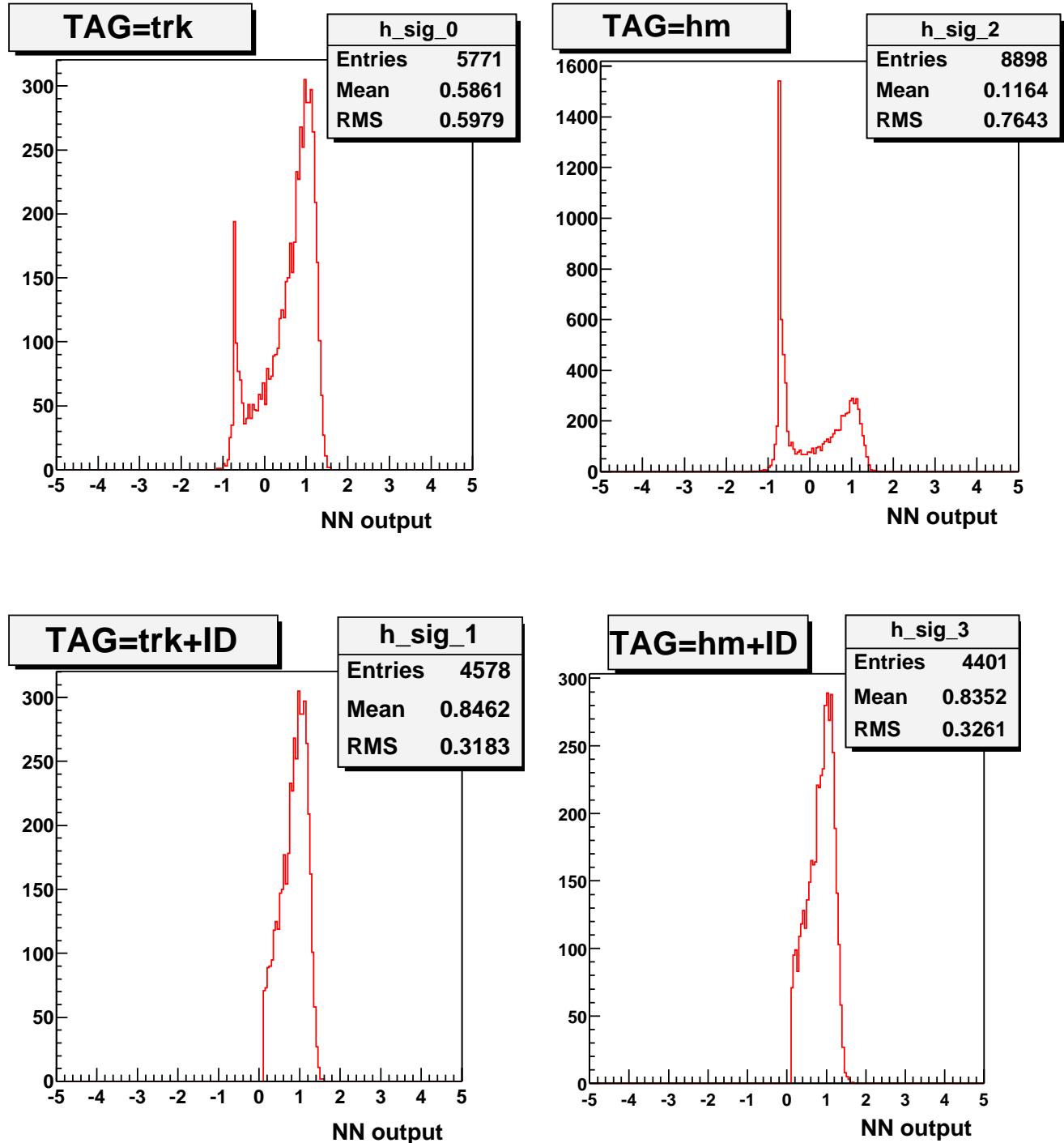


Figure 67: NN output for the probe object corresponding to the CC EMID efficiency measurement. EMID in CC = Neural Net output > 0.1 . Training was done with 800 epochs. 7 NN variables are used: those in HMx7 minus Z(primary vertex) (EM3 rphi-width, four EM floor energy fractions, log(E)) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : chi^2 < 5$) is applied to the tag object. Top plots shows the denominator NN output i.e. when no cut is applied to the probe object. Bottom plots show numerator NN output when EMID cut is applied to the probe object.

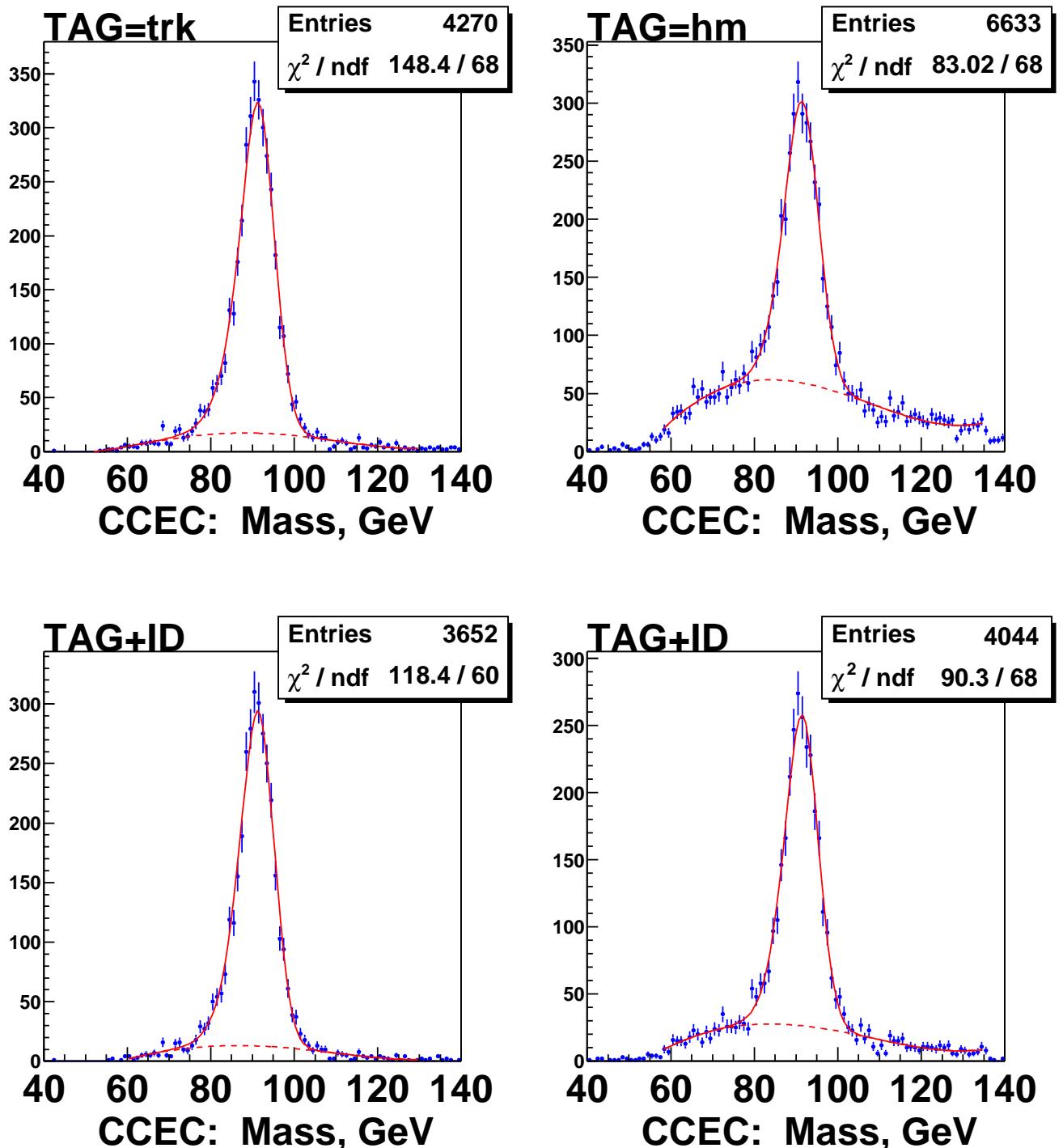


Figure 68: CCEC diEM mass distributions for the EC EMID efficiency measurement. EMID in EC = Neural Net output > -0.1 . Training was done with 500 epochs. 8 NN variables are used: those in HMx8 minus Z(primary vertex) (EM3 rphi-width, four EM floor energy fractions, $\log(E)$) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : chi^2 < 5$) is applied to the tag object. Top plots show the denominator mass distributions i.e. when a cut is applied to the tag object only. Bottom plots show numerator mass distributions i.e. when EMID cut is applied to the probe object.

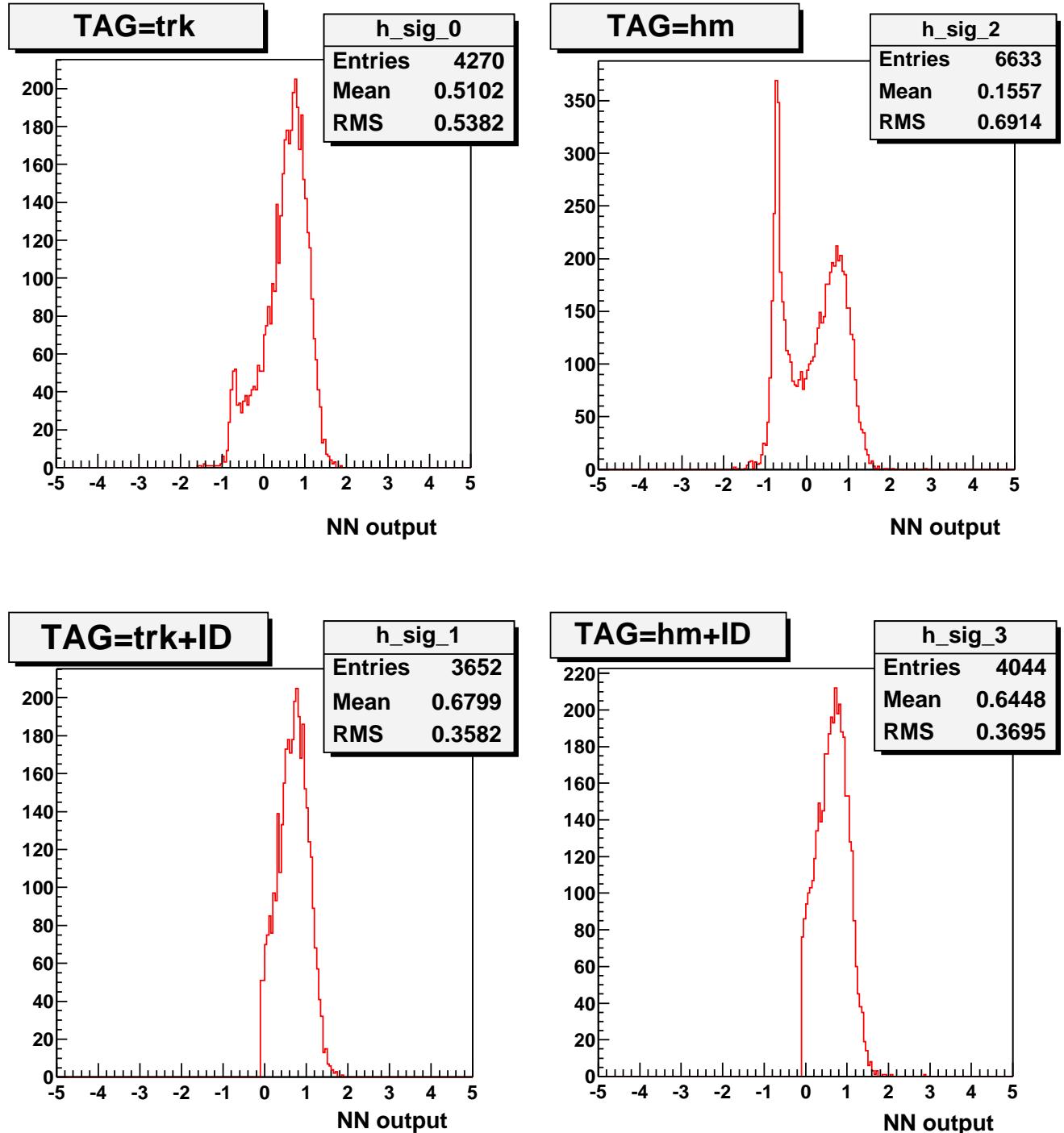


Figure 69: NN output for the probe object corresponding to the EC EMID efficiency measurement. EMID in EC = Neural Net output > -0.1 . Training was done with 500 epochs. 8 NN variables are used: those in HMx8 minus Z(primary vertex) (EM3 $rphi$ -width, four EM floor energy fractions, $\log(E)$) plus track isolation. Left column corresponds to the case when the tag object is required to have a track match, whereas in the right column the calorimeter based cut ($mboxHM8 : chi^2 < 5$) is applied to the tag object. Top plots shows the denominator NN output i.e. when no cut is applied to the probe object. Bottom plots show numerator NN output $4\bar{e}$. when EMID cut is applied to the probe object.

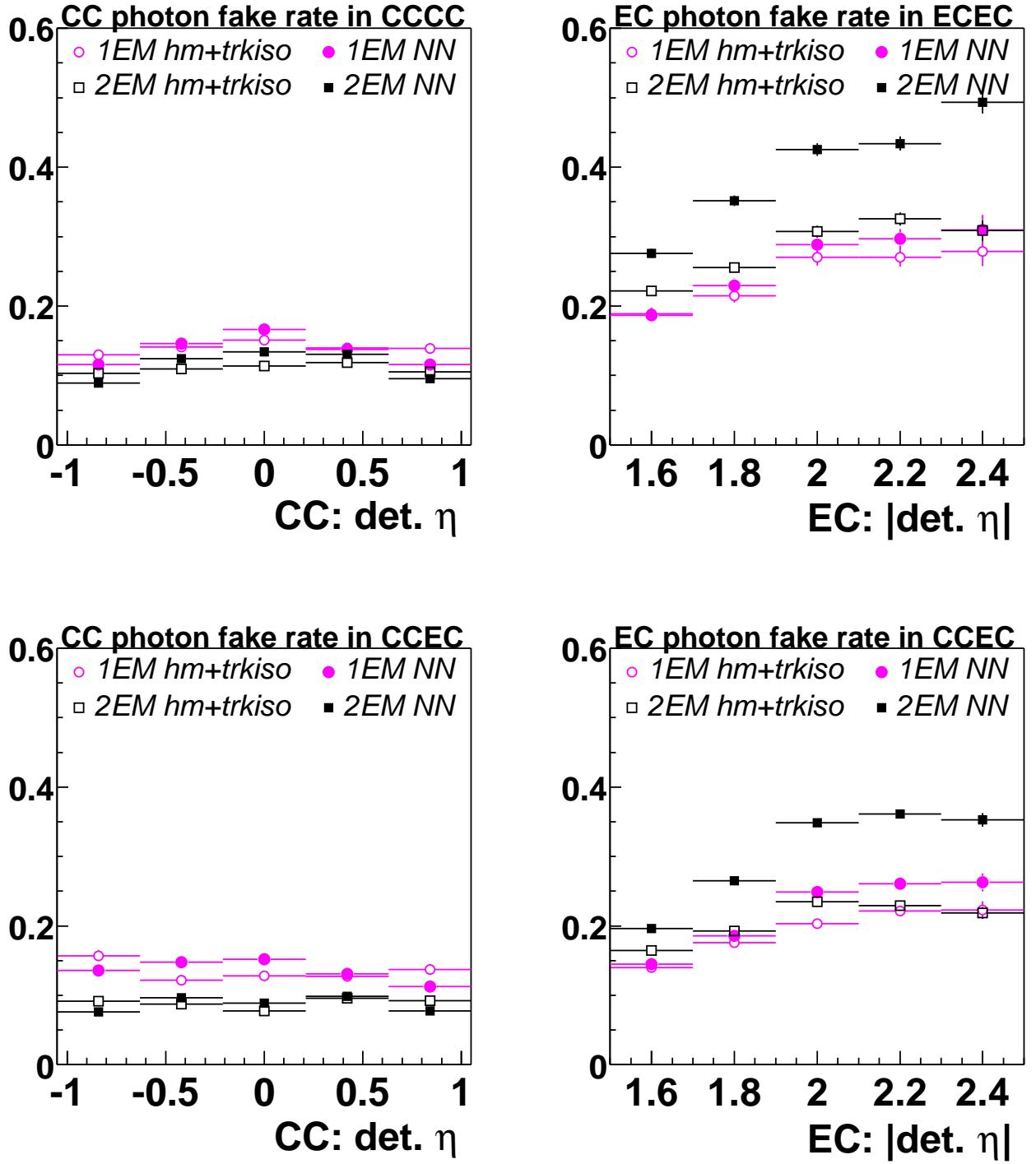


Figure 70: Photon fake rates vs. detector η . Top left: CC rate in CCCC subset. Bottom left: CC rate in CCEC subset. Top right: EC rate in ECEC subset. Bottom right: EC rate in CCEC subset. Circles show fake rate measured in Single EM sample. Squares show fake rate measured in diEM sample (with invariant mass cut of $M!=[70,100]$ GeV). Open markers correspond to HMATRIX + track isolation ID. Solid markers correspond to NN. NN trained with 1500 epochs (on the plateau).

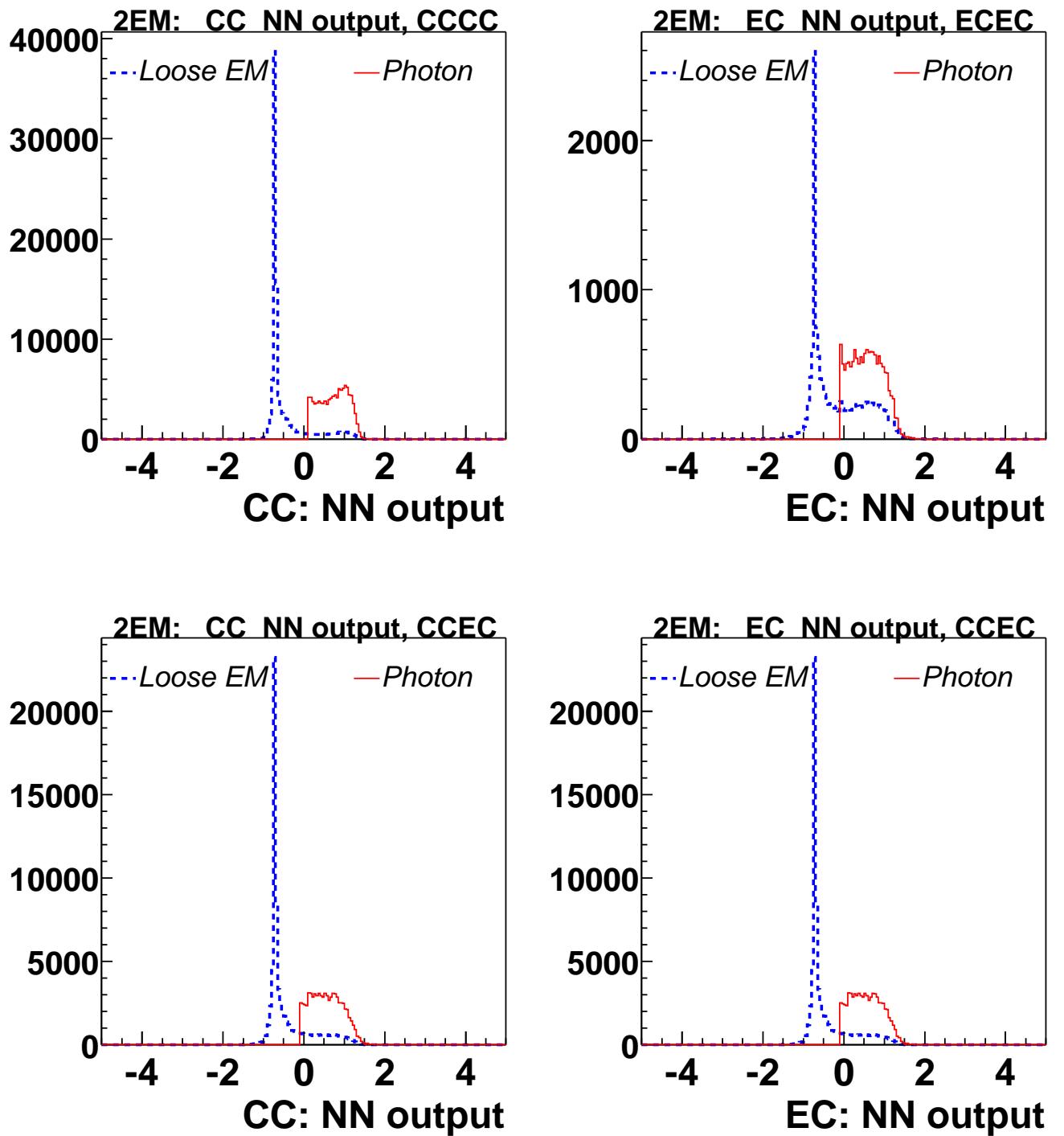


Figure 71: NN output in 2EM sample for the photons (photon fake rate numerator) and loose em objects. NN trained with 1500 epochs (on the plateau).

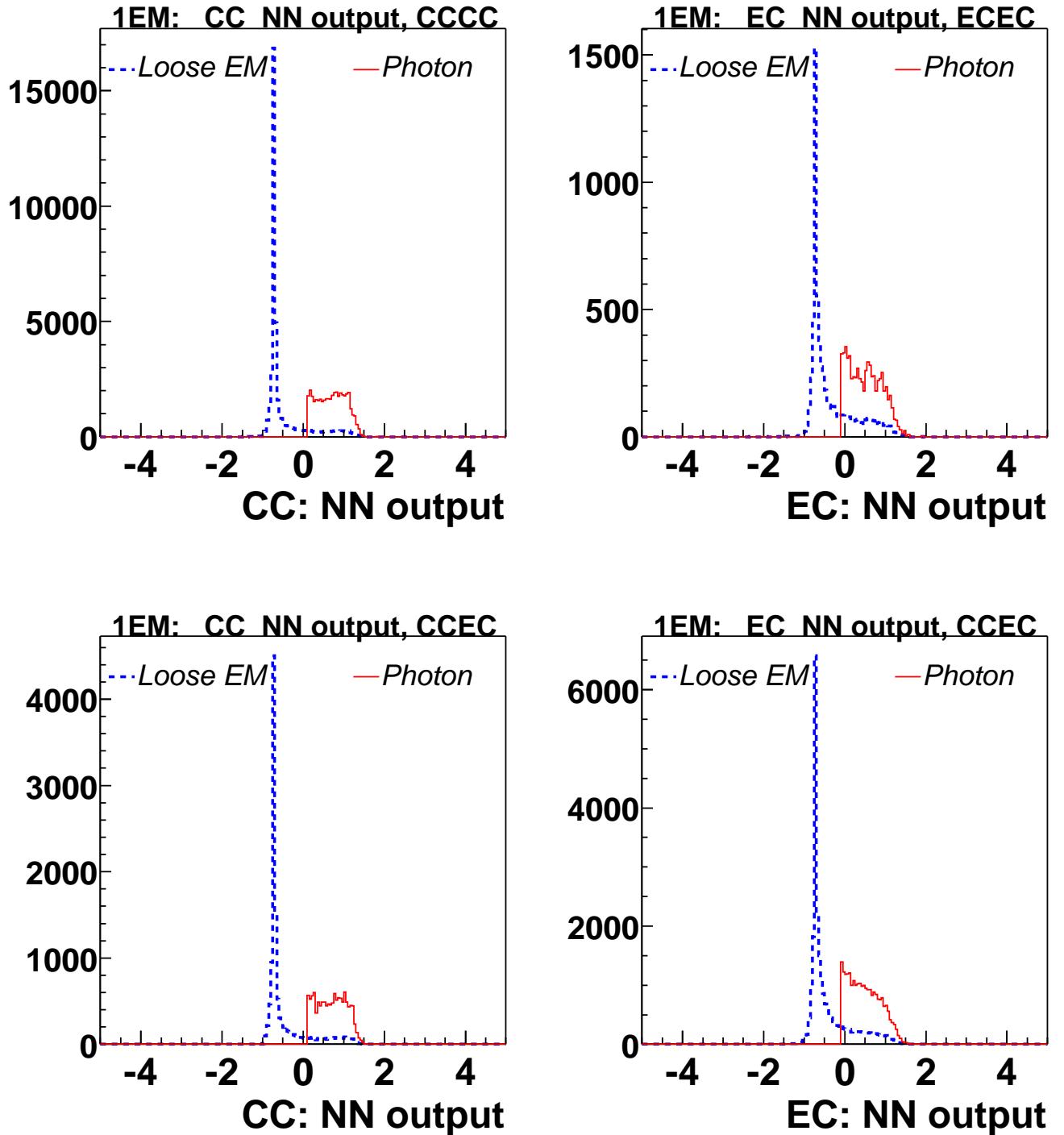


Figure 72: NN output in 1EM sample for the photons (photon fake rate numerator) and loose em objects. NN trained with 1500 epochs (on the plateau).

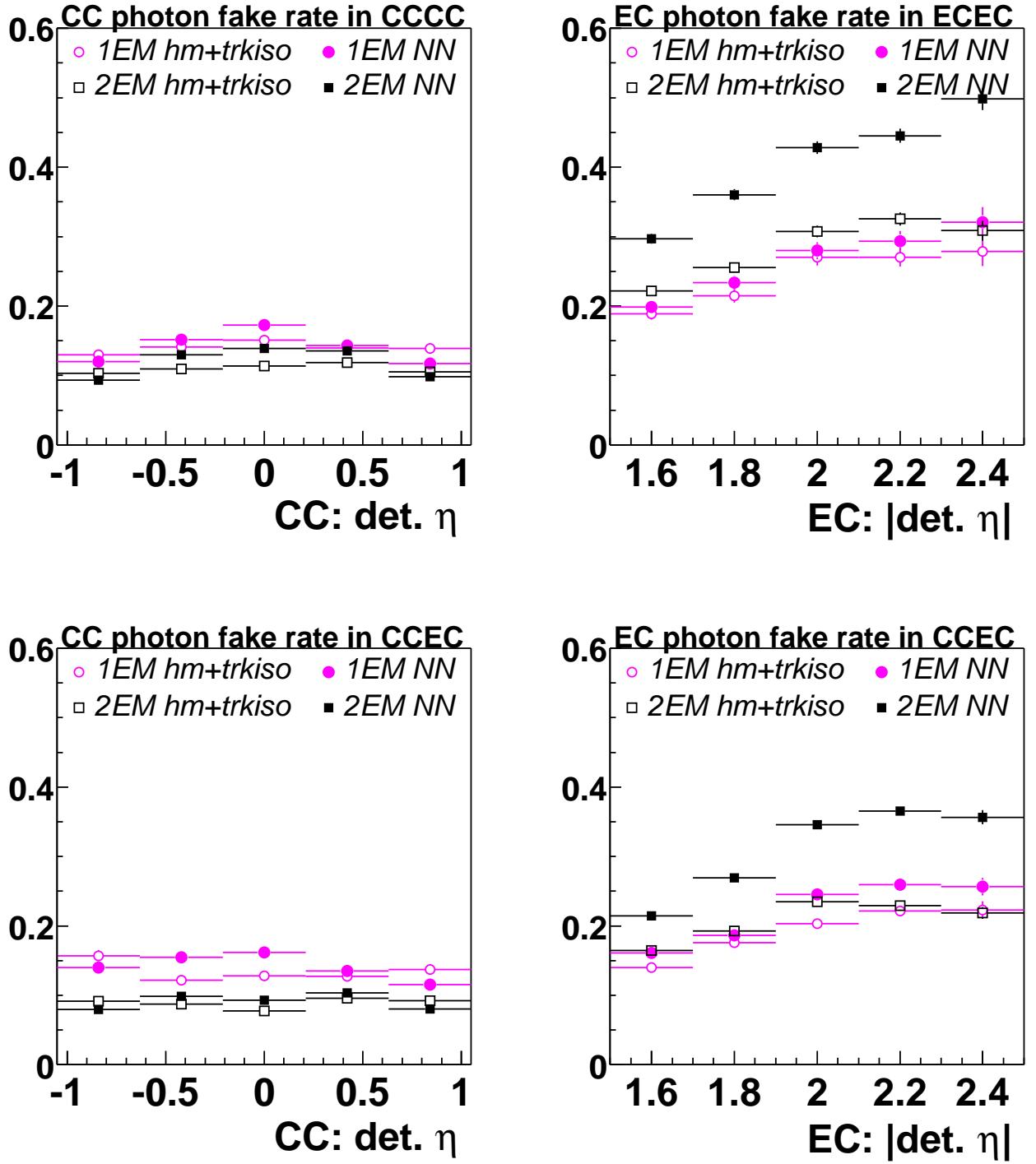


Figure 73: Photon fake rates vs. detector *eta*. Top left: CC rate in CCCC subset. Bottom left: CC rate in CCEC subset. Top right: EC rate in ECEC subset. Bottom right: EC rate in CCEC subset. Circles show fake rate measured in Single EM sample. Squares show fake rate measured in diEM sample (with invariant mass cut of $M!=[70,100]$ GeV). Open markers correspond to HMATRIX + track isolation ID. Solid markers correspond to NN. NN trained with 800 (CC) and 500 (EC) epochs (before the plateau).

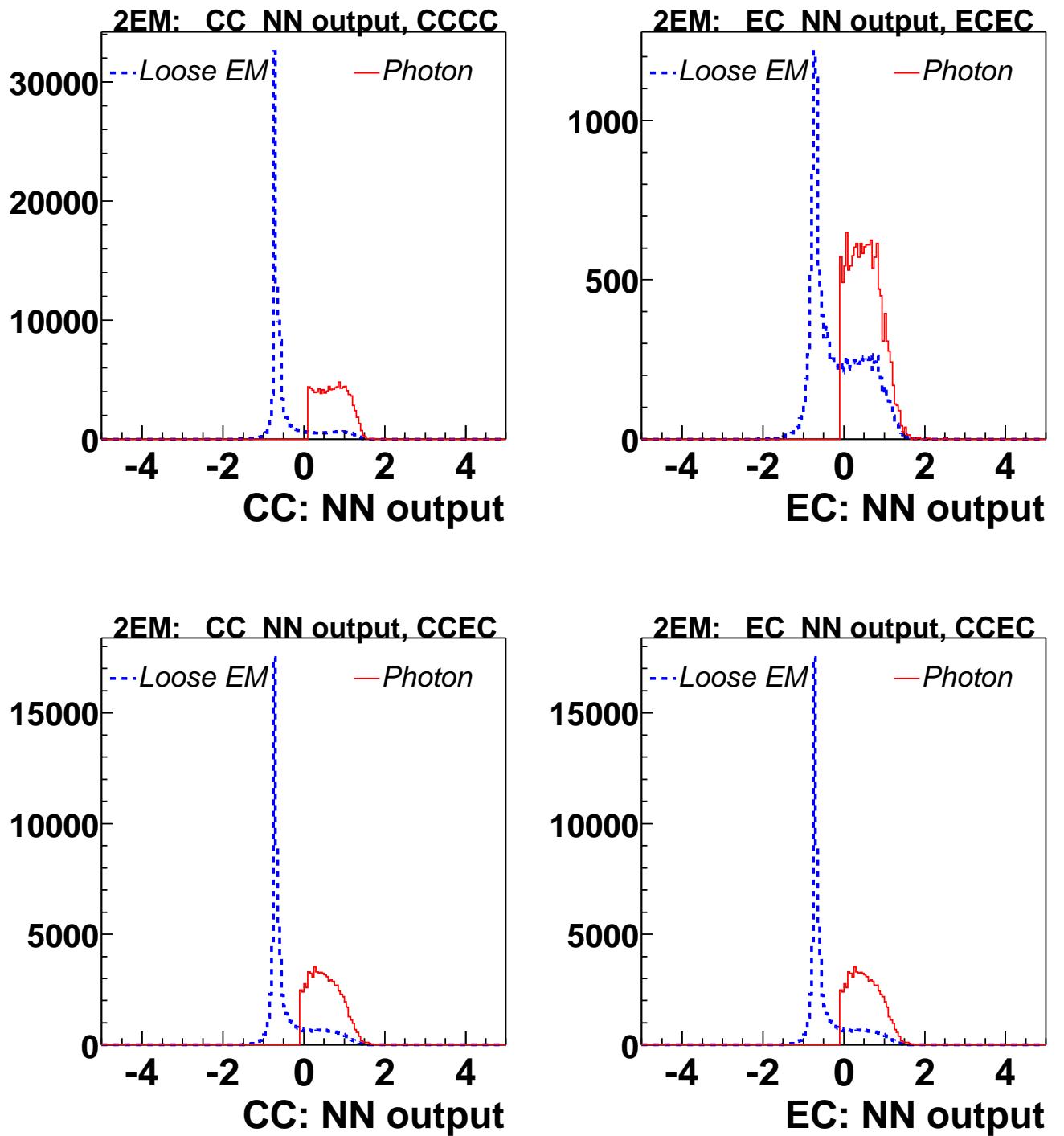


Figure 74: NN output in 2EM sample for the photons (photon fake rate numerator) and loose em objects. NN trained with 800 (CC) and 500 (EC) epochs (before the plateau).

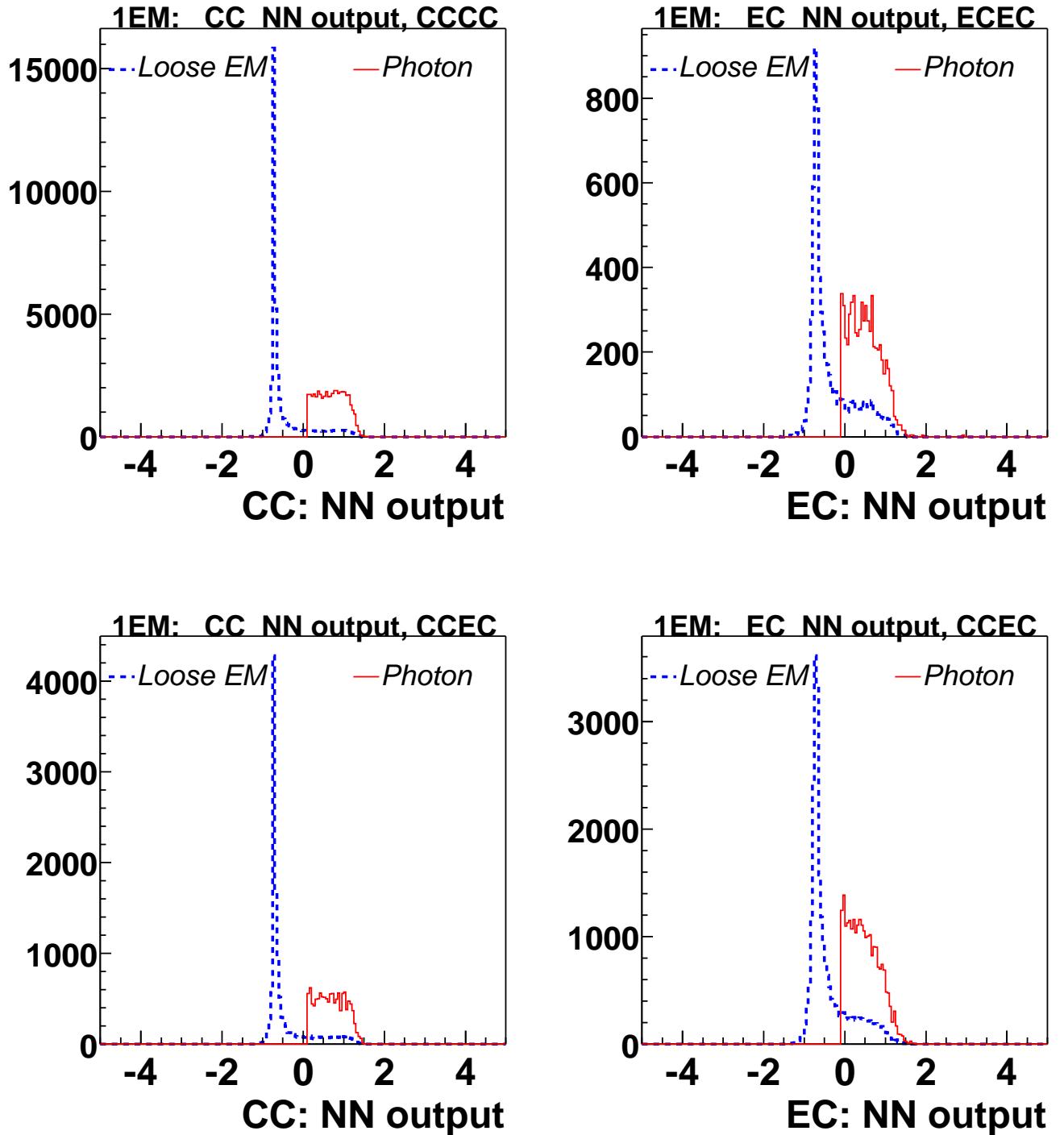


Figure 75: NN output in 1EM sample for the photons (photon fake rate numerator) and loose em objects. NN trained with 800 (CC) and 500 (EC) epochs (before the plateau).

References

- [1] A. Melnitchouk, “Search for non-SM Light Higgs Boson in the $h \rightarrow \gamma\gamma$ Channel at DØ in Run II”, DØ note 4351, 2004.
- [2] A. Melnitchouk, “Search for non-SM Light Higgs Boson in the $h \rightarrow \gamma\gamma$ Channel at DØ in Run II”, DØ note 4374, 2004.
- [3] http://www-d0.fnal.gov/home/~melnit/WINTER_2004/June17_h.ps,
Higgs-dilepton group meeting report.
- [4] G. Landsberg, K. Matchev, Phys. Rev. **D62**, 035004 (2000).
- [5] B. Abbot *et al.* (DØ Collaboration), Phys Rev. Lett. **82**, 2244 (1999);
- [6] Drew Alton, Andrew Askew, H.T. Diehl, Yurii Maravin, Sean Mattingly, “Measurement of W Gamma Events in DØ Run II Data”, DØ note 4410, 2004.
- [7] Brian Lauer, Ph.D Thesis, Iowa State University (Unpublished) (1997).
- [8] Jan Stark, private communication.
- [9] <http://schwind.home.cern.ch/schwind/MLPfit.html>
- [10] <http://root.cern.ch/root/html/TMultiLayerPerceptron.html>
- [11] Andrew Askew, Ms.Sc. Thesis, Rice University (Unpublished) (2001); Andrew Askew, private communication.
- [12] Yurii Maravin, private communication.